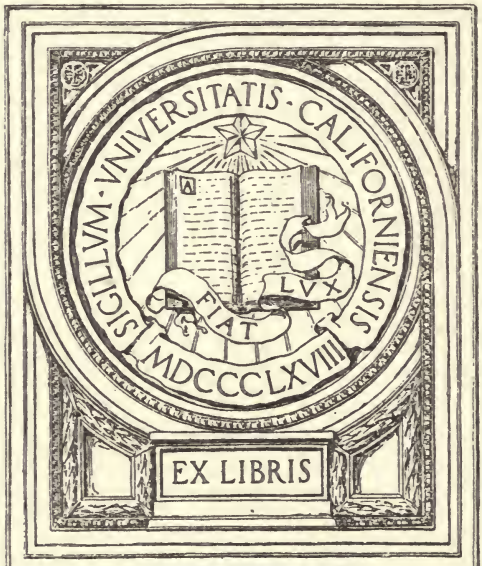
LONGMANS'
MODERN MATHEMATICAL SERIES

THE TEACHING OF ALGEBRA

(INCLUDING TRIGONOMETRY)

T. PERCY NUNN

GIFT OF
Miss Emily Palmer



EDUCATION DEPT.

16 2411
700
LANGE LIBRARY OF EDUCATION
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA

Emily G. Palmer

Digitized by the Internet Archive
in 2008 with funding from
Microsoft Corporation

http://www.archive.org/details/teachingofalgebr00nunnrich

LONGMANS' MODERN MATHEMATICAL SERIES

General Editors

P. ABBOTT, B.A., C. S. JACKSON, M.A.
F. S. MACAULAY, M.A., D.Sc.

THE TEACHING OF ALGEBRA

BY THE SAME AUTHOR

UNIFORM WITH THIS VOLUME.

EXERCISES IN ALGEBRA (INCLUDING TRIGONOMETRY). Part I. Without Answers, 3s. 6d.; with Answers, 4s.

CONTENTS.—Section I, "Non-directed Numbers"; Section II, "Directed Numbers"; Section III, "Logarithms"; Supplementary Exercises.

EXERCISES IN ALGEBRA (INCLUDING TRIGONOMETRY). Part II. Without Answers, 6s.; with Answers, 6s. 6d.

CONTENTS.—Section IV, "Mainly Revision"; Section V, "The Trigonometry of the Sphere"; Section VI, "Complex Numbers"; Section VII, "Periodic Functions"; Section VIII, "Limits"; Section IX, "Statistics".

LONGMANS, GREEN AND CO.

LONDON, NEW YORK, BOMBAY, CALCUTTA, AND MADRAS

Longmans' Modern Mathematical Series

THE TEACHING OF ALGEBRA

(INCLUDING TRIGONOMETRY)

BY

T. PERCY NUNN, M.A., D.Sc.

PROFESSOR OF EDUCATION IN THE UNIVERSITY OF LONDON
VICE-PRINCIPAL OF THE L.C.C. LONDON DAY TRAINING COLLEGE (UNIVERSITY OF LONDON); FORMERLY SENIOR MATHEMATICAL AND SCIENCE MASTER WILLIAM ELLIS SCHOOL

WITH DIAGRAMS

LONGMANS, GREEN AND CO.
39 PATERNOSTER ROW, LONDON
FOURTH AVENUE & 30TH STREET, NEW YORK
BOMBAY, CALCUTTA, AND MADRAS
1914

LB1645
N8
Educ.
Dept.

Gift of Miss Emily Palmer
to Educ. Dept.

PREFACE.

In 1909 and 1910 the author had the honour of giving courses of lectures on the teaching of Algebra addressed respectively to masters and mistresses in Secondary Schools. The present volume is a practical handbook based upon those lectures, and containing what seem to the author the most useful things he has learnt during the fifteen years of his work as a mathematical master and the ten years in which it has been his duty to discuss with teachers, actual and prospective, the problems of their craft. It is accompanied by two collections of examples—"Exercises in Algebra, Part I," and "Exercises in Algebra, Part II"—which are intended together to cover all stages of school instruction in the subject. Thus the three volumes constitute a single work. In the view of the author the term "Algebra" should include in its reference all the Trigonometry, plane and spherical, which it is desirable to teach in schools, together with an exposition of the fundamentals of the Calculus. He has sought, therefore, to present these subjects, both in this book and in the "Exercises," as a unified whole.

"Exercises, Part I," is intended to supply materials

632257

for a course which every boy or girl who remains at a secondary school until the age of sixteen or seventeen may reasonably be expected to cover. Thus it meets and in some directions exceeds the present requirements of University Entrance and similar examinations. An important note upon this point will be found on page 60 of the present book. "Exercises, Part II," presents a continuation of this universal course, to be taken, as a whole or in part, by boys and girls who remain at school until the age of eighteen or nineteen, and by those who, at an earlier age, forge ahead of their fellows. It is hoped that it may also prove useful to students preparing to enter the scientific professions and to students in Training Colleges for Teachers.

In "Exercises, Part I," explanatory matter has been limited to a few brief notes—on the ground that in the earlier stages of the pupil's progress exposition is of little use unless given verbally by the teacher. Thus the discussions and arguments presupposed in the exercises have been relegated to the present volume, where they could be given in a form which, it is hoped, the younger teacher will find more helpful, and his experienced colleague more suggestive and provocative. In "Exercises, Part II," each set of examples is accompanied by exposition intended for the student's reading. To make it suitable for this purpose it has seemed necessary to abandon the traditional reticence of the mathematical textbook—a fact to be borne in mind in connexion with the otherwise misleading length of the book. The corresponding sections of the present book consist of critical introductions, commentaries upon the exposition and exercises, suggestions for illustrations

and other teaching devices, and solutions of typical examples.

There is considerable evidence that teachers of the present generation are no longer satisfied with either the curriculum or the methods of instruction in Algebra which they have inherited from their predecessors. Moreover, there are signs that their dissatisfaction is reflected in the public examining bodies which exercise such immense influence upon mathematical teaching in this country. These circumstances may be held to justify an addition to the few works written expressly for the teacher who wishes, while maintaining his hold upon all that is sound in the traditional methods, to orient himself in accordance with the present drift of well-considered opinion. The author has worked out in "Exercises, Part I," a curriculum congruent on the whole with the recent Report of the Committee of the Mathematical Association, though in some respects more radical and in some more conservative. He has noted with special satisfaction that, taking the whole work into consideration, the programme of studies which he has proposed is in close agreement with the one outlined by Dr. A. N. Whitehead in the very valuable address printed in the *Mathematical Gazette* for March, 1913. With regard to method, the author has sought to take due account both of the pedagogical pragmatism of which Professor Perry has long been our most influential advocate and also of the modern critical movement represented so brilliantly in this country by the authors of *Principia Mathematica*. In addition, recognizing that mathematical ideas are

apt to become "flat" through long confinement in textbooks, he has sought, by drawing them afresh from their historical sources, to present them with as much as possible of their original vigour.

Acknowledgment has been made in the text wherever the author has consciously taken advantage of the work of writers whose ideas have not yet been absorbed into the common stock. He has the pleasure of recording here many other obligations, general and specific. From his three Editors he has constantly received criticism and advice of the most helpful kind. Mr. C. O. Tuckey of Charterhouse read a good deal of the earlier part of the work in typoscript and made several important suggestions, gratefully accepted. Professor C. Spearman of University College very kindly read and criticised the section on Statistics—a subject to which he has himself made such important contributions. Similar service was rendered, in connexion with other sections, by Messrs. G. B. Jeffery of University College, B. M. Neville of the William Ellis School, and E. Wasserman of Owen's School. Dr. L. Silberstein, from his inexhaustible store of fertile mathematical ideas, has kindly contributed those embodied in several important examples. Intercourse with Mr. Benchara Branford and Mr. David Mair has had an influence upon the book which is not the less substantial because it is not located in any particular pages. The author's thanks are due also to Miss Doris Brookes, Miss Elsa Nunn and Miss Clotilde von Wyss for assistance given in the preparation of the diagrams, and to a number of his present and former students for undertaking the ungrateful task of providing the answers to the examples.

Lastly, he has to acknowledge the help received at every stage of the work from his colleague, Miss Margaret Punnett, who not only drew all the diagrams which required special care or laborious calculations, corrected all the proofs, and compiled the index of the present volume, but also gave constant criticism and encouragement of the greatest value.

It should be unnecessary to add that the author does not seek by these acknowledgments to make his friends share the responsibility for anything in the work that may be judged unsatisfactory, or to suggest that they accept his views upon all the disputed or disputable questions to which he has offered answers.

London Day Training College
(University of London),
May, 1914.

CONTENTS.

GENERAL INTRODUCTION.

CHAPTER PAGE

I. THE NATURE OF ALGEBRA 1

II. METHOD AND CURRICULUM 16

III. THE FORMULA 25

IV. THE GRAPH 31

PART I.

ALTERNATIVE SCHEMES OF STUDY 50

V. INTRODUCTION TO PART I 51

SECTION I.

NON-DIRECTED NUMBERS.

THE EXERCISES OF SECTION I 62

VI. THE PROGRAMME OF SECTION I (Exercises I-XVI) . 63

VII. FACTORIZATION 82

A. Factorization of $ac + bc$ (p. 82); B. Factorization of $a^2 - b^2$ (p. 87).

VIII. SQUARE ROOT. SURDS 90

A. The Calculation of Square Roots (p. 90); B. The Radical Form (p. 93).

IX. FRACTIONS 96

A. Fractions with Monomial Denominators (p. 96); B. Fractions with Binomial Denominators (p. 98).

X. CHANGING THE SUBJECT OF A FORMULA . . . 104

XI. PROGRAMME OF SECTION I (Exercises XVII-XXVI) . 109

CHAPTER PAGE

XII. Direct Proportion 117

XIII. Trigonometrical Ratios (I) 121

A. The Tangent of an Angle (p. 121); B. The Sine and Cosine, Vectors (p. 124).

XIV. Trigonometrical Ratios (II) 129

A. Circles of Latitude, Middle Latitude Sailing (p. 129); B. Relations between the Sine, Cosine, and Tangent (p. 132).

XV. The Combining of Formulæ 136

A. The Determination of Constants in a Formula (p. 136); B. Common Values of Two Relations (p. 139); C. Elimination (p. 142).

XVI. Further Types of Proportionality . . . 145

A. Inverse Proportion (p. 145); B. Direct Proportion to the Square or Square Root (p. 149); C. Inverse Proportion to the Square or Square Root (p. 152); D. Combinations of Types of Proportion (p. 154).

SECTION II.

DIRECTED NUMBERS.

The Exercises of Section II 157

XVII. The Programme of Section II (Exercises XXVII-XXXVIII) 159

XVIII. Directed Numbers 181

A. The Uses of Directed Numbers (p. 181); B. Algebraic Addition and Subtraction (p. 184); C. The Multiplication and Division of Directed Numbers (p. 193).

XIX. Constant-Difference Series 199

A. The Summation of Constant-Difference Series (p. 199); B. The Calculation of Certain Areas and Volumes (p. 203).

XX. Algebraic Multiplication 207

A. Algebraic Multiplication (p. 207); B. The Binomial Expansion (p. 211).

CHAPTER PAGE

XXI. Positive and Negative Indices 214
A. The Uses and Laws of Positive Indices (p. 214); B. Negative Indices (p. 217).

XXII. Algebraic Division 222
A. Algebraic Division (p. 222); B. Geometric Series (p. 224).

XXIII. The Complete Number-Scale 228

XXIV. The Programme of Section II (Exercises XXXIX-L) 235

XXV. Linear Functions. Extended Use of Sine, Cosine, and Tangent 258
A. Linear Functions (p. 258); B. Extension of meaning of Sine and Cosine (p. 261).

XXVI. The Hyperbolic and Parabolic Functions . . 264
A. Hyperbolic Functions (p. 264); B. Parabolic Functions (p. 266); C. Quadratic Equations (p. 270); D. Inverse Functions (p. 274).

XXVII. Wallis's Law 279
A. Area Functions (p. 279); B. Differential Formulæ (p. 282).

XXVIII. The Calculation of π and the Sine-Table . . 292
A. The Calculation of π (p. 294); B. The Calculation of the Sine-Table (p. 295).

SECTION III.

LOGARITHMS.

The Exercises of Section III 298

XXIX. The Programme of Section III 299

XXX. The Graphic Solution of Growth-Problems . 312

XXXI. The Gunter Scale 319

XXXII. Logarithms 325
A. The Slide Rule (p. 325); B. Logarithms (p. 329).

XXXIII. Common Logarithms 333
A. Gunter's Scale and Logarithms Obtained by Calculation (p. 333); B. Common Logarithms (p. 335); C. The Use of Tables (p. 337).

CHAPTER PAGE
XXXIV. The Logarithmic and Antilogarithmic Functions 341
XXXV. Nominal and Effective Growth-Factors . . 346

EXERCISES SUPPLEMENTARY TO SECTIONS II AND III.

Supplementary Exercises 354
XXXVI. The Programme of Exercises LX-LXV . . 355
XXXVII. The Programme of Exercises LXVI-LXIX . . 366

PART II.

Alternative Schemes of Study 380
XXXVIII. Introduction to Part II 381

SECTION IV.

MAINLY REVISION.

The Exercises of Section IV 402
XXXIX. Number-Systems and Numerical Operations . 403
XL. Functions 422
XLI. The Exponential Function and Curve . . . 428

SECTION V.

THE TRIGONOMETRY OF THE SPHERE.

The Exercises of Section V 440
XLII. Projections 441
XLIII. The Trigonometry of Spherical Triangles . . 454

SECTION VI.

COMPLEX NUMBERS.

The Exercises of Section VI 468
XLIV. The Nature of Complex Numbers 469
XLV. Relations between a Real and a Complex Variable 478
XLVI. Relations between two Complex Variables . 486

SECTION VII.

PERIODIC FUNCTIONS.

CHAPTER		PAGE
	THE EXERCISES OF SECTION VII	498
XLVII.	THE CIRCULAR FUNCTIONS	499
XLVIII.	WAVE-MOTION	515
XLIX.	DIFFERENTIAL FORMULÆ OF THE CIRCULAR FUNCTIONS	528
L.	THE HYPERBOLIC FUNCTIONS	531

SECTION VIII.

LIMITS.

	THE EXERCISES OF SECTION VIII	540
LI.	DIFFERENTIATION AND INTEGRATION	541
LII.	EXPANSIONS. SUPPLEMENTARY EXAMPLES	557

SECTION IX.

STATISTICS.

	THE EXERCISES OF SECTION IX	568
LIII.	FREQUENCY-DISTRIBUTION	569
LIV.	THE CALCULATION OF FREQUENCIES. PROBABILITY	583
LV.	CORRELATION	602
	INDEX	611

COLOURED DIAGRAMS.

Figs. 48, 49, 50	*To face page*	200
Fig. 51	,, ,,	201
Fig. 58	,, ,,	208
Fig. 59	,, ,,	209
Figs. 60, 61, 62	,, ,,	210

GENERAL INTRODUCTION.

CHAPTER I.

THE NATURE OF ALGEBRA.

§ 1. A discussion of the proper definition of "Algebra" may easily be carried to the point where its interest becomes academic rather than practical. It will be limited here to a review of the general nature of the topics traditionally studied and taught under a name of ancient and obscure origin.[1] Inspection of these topics shows the presence of at least four important elements: Analysis, the direct use of symbolism, the extended use of symbolism, and the manipulation of symbolism.

§ 2. *Analysis.*—The most fundamental is analysis. The difficulty of finding a precise boundary between arithmetic and algebra (as these terms are commonly understood) is well known. It is due to the fact that the distinction between them consists not so much in a difference of subject-matter as in a difference of attitude towards the same subject-matter. A simple example may make the difference clear. A small boy has learnt that the "area" of a figure is the number of unit squares (say, square inches) which would entirely cover it. With this definition before him he is asked to determine the area of a rectangle measuring 7 inches by 5. He soon observes that the unit squares into which the figure is to be mapped out can be regarded as forming five rows each con-

[1] It is unfortunate that a writer is compelled by convention to place his introductory chapter at the threshold of his work. The disadvantages of the arrangement are obvious. An introduction often represents the author's attempt to meet difficulties and to answer objections which the reader has not yet felt. It is therefore liable to be misunderstood or misused. It is to be hoped that the reader who finds in the present chapter little relevance to the urgent problems of algebra teaching will by judicious skipping reach the end without serious waste of time. He will find at many points of the sequel references which will send him back, perhaps to a more profitable reading of these preliminary discussions.

taining seven squares. This observation enables him to shorten the process of finding the area; for it is obvious that the rectangle must contain $7 \times 5 = 35$ square inches. So far arithmetic. But now let the boy's attention shift from the actual manipulation of the figures to the *process* which the manipulation follows; and let him observe that the essence of that process is the multiplication of the length of the rectangle by its breadth. At this moment he has crossed the frontier which separates arithmetic from algebra; for it is an important part of the business of algebra to disengage the essential features of an arithmetical process of given type from the numerical setting which a particular case presents.

The result of the boy's analysis of the process of area-calculation will be expressed in a statement or "rule" that makes no reference to the particular numbers, 7 and 5, and would therefore hold good in any other area-calculation of the same type. For this reason it is usually called a "generalization". Strictly speaking, however, the terms analysis and generalization refer to two distinct mental movements. In the former I bring to light the essential process concealed in a particular or accidental numerical garb. In the latter I recognize that this process may be followed identically in solving all problems of the same type. The distinction is not a trivial one, but demands the teacher's serious attention. The neglect of it is largely responsible for the common belief that the process of generalization must of necessity build upon numerous examples of the truth to be generalized, and that the security of the result depends upon the number of instances upon which it is based. This belief makes the schoolmaster regard "generalizing from a single instance" as one of the most dangerous manifestations of original intellectual sin.

The truth is that in some cases the certainty of a generalization does depend upon the multiplicity of its data, while in other cases the number of data is logically irrelevant; a single instance will establish the rule as securely as a hundred. Thus, when John Wallis (1655) had shown that what we now call the "integrals" of x, x^2, x^3, x^4 are $\frac{1}{2}x^2$, $\frac{1}{3}x^3$, $\frac{1}{4}x^4$, $\frac{1}{5}x^5$, he felt entitled to generalize the result and to assume that the integral of x^{n-1} would be $\frac{1}{n}x^n$ for all integral values of n. This was a generalization of the former kind and had probable

truth only. The degree of probability of such a generalization clearly depends in the first instance upon the number of data upon which it is based. Its credit will subsequently rise if results deduced from it are found to agree with known truths, or will collapse if a "negative instance" can be produced. Thus the numbers 41, 43, 47, 53, 61, 71, 83, 97, 113, 131, etc., are all prime and are all included in the formula $n^2 + n + 41$. But the hypothesis that this formula always yields primes is at once discredited when it is seen that it fails for $n = 40$.[1] On the other hand there is no possibility that the rule for the area of a rectangle should break down.[2] Its certainty is due, not to verification in a large number of instances, nor to its simplicity, but merely to the fact that it is founded on analysis.

It may be objected that the two types of generalization cannot be distinguished in this way; that, for example, Wallis had to analyse his individual results before he could include them all in a single algebraic statement. This is, of course, true, but it should be noted that in Wallis's case the analysis was limited to the discovery that, in several instances, expressions of the form x^{n-1} actually have integrals of the form $\frac{1}{n}x^n$. Analysis in the sense intended here would deal essentially with *the circumstances on which this relation depends.* An analysis in this sense of any one of Wallis's instances of integration would establish for ever the certainty of all instances covered by his formula.[3] Thus the question whether a generalization requires the evidence of a number of examples is the question whether or not it is founded on analysis. If it is not so founded the generalization has at best only probability in its favour, a probability which increases with the number of instances in which it is verified. If it is so founded, though in practice several examples may be needed to point the way to the generalization, yet in theory a single instance is sufficient to render it certain.[4]

[1] Quoted from Jevons, *Principles of Science.*

[2] It is assumed for the present that the dimensions are integral. See § 3.

[3] I.e. in which n is integral.

[4] Logicians (e.g. Bradley) have called the first kind of generalization an "empirical universal," the second kind an "unconditional universal". The philosophically minded reader may follow

Each form of generalization has played an important part in the history of mathematical discovery. Each has its place in school instruction. They have, however, very unequal value. While a mathematical truth may for a time be accepted and used—as Newton accepted and used the binomial theorem for a fractional index—on the guarantee of a number of instances in which it is known to hold good, yet its position is not felt to be satisfactory until it has been placed on the firmer foundations of analysis. On the other hand, it must not be forgotten that the power to reach by analysis a complete guarantee of a mathematical truth is subject to the weakness which limits all human powers. Thus Pythagoras —probably as the result of the examination of a number of special instances—first enunciated the proposition that the sum of the squares on the sides of a right-angled triangle is equal to the square on the hypothenuse. Euclid's "demonstration" was an attempt to guarantee the truth of the proposition by analysis. Apparently the analysis seemed to Euclid complete, and satisfied geometers down to modern times. But it involves the proposition about congruent triangles known to our youth as Prop. 4, and this truth is not really guaranteed by Euclid's analysis of it. A recent brilliant critic, generalizing illegitimately, like Macaulay, from the single instance of his own amazing precocity, says that Euclid's proof of Prop. 4 "strikes every intelligent child as a juggle". Euclid's proof of the theorem of Pythagoras thus offers an instance of a piece of mathematical analysis, long thought to be adequate, which the modern critical sense requires to be corrected and made more complete. Similar incidents are constantly occurring in various parts of mathematical theory. They show us that mathematical reasonings do not necessarily issue from the brains of mathematicians in full-blown perfection, like Athene from the head of Zeus. For them, as for all other human productions, perfection is a goal to be reached, if at all, only after a long process of purification by criticism and reconstruction.

§ 3. *Direct Use of Symbolism.*—For the task of algebra as described in the foregoing section only two tools are, in

up the question raised above in James's *Principles of Psychology*, i., ch. XII., and Prof. Bosanquet's paper in the *Proceedings of the Aristotelian Society*, 1910-11.

principle, necessary: the power (which every mind possesses in some measure) of discerning the abstract essential process in the concrete arithmetical case, and a sufficient command of language to express it when discerned. But in practice something more is needed. Neither the analysis of arithmetical procedure nor the expression of the results of such analysis can proceed very far without the help of symbolism. Thus it is difficult to suppose that the generalization commonly expressed in the form

$$(x + a)^n = x^n + nx^{n-1}a + \frac{n(n-1)}{\underline{|2}}x^{n-2}a^2 + \ldots + a^n$$

could ever have been reached without the aid of symbols. Moreover its expression in language would not only be intolerably prolix but almost useless for practical purposes; only the conciseness of the symbolic expression makes it possible for the student either to grasp the generalization or to apply it. The development of a symbolism with the properties of making analysis easier and the expression of its results more concise and available is, then, the second fundamental element in algebra.

This function of symbolism has an importance that reaches much beyond the field of school mathematics. "The ideal of mathematics," writes Dr. Whitehead,[1] "should be to erect a calculus to facilitate reasoning in connexion with every province of thought, or of external experience, in which the succession of thoughts or of events can be definitely ascertained and precisely stated. So that all serious thought which is not philosophy, or inductive reasoning, or imaginative literature, shall be mathematics developed by means of a calculus." The definition of algebra suggested by this passage is, of course, much too wide for our present purpose, but it brings out a point of great importance. The algebra with which we are all familiar is only one of an indefinite number of possible algebras. Wherever there is a field for inquiry of a certain type an algebra may be invented to facilitate that inquiry.

The field of common algebra is that of numbers and their relations, but the school curriculum itself yields an example of an algebra concerned with an entirely different region. For in chemical formulæ and "equations" we have a system

[1] *Universal Algebra*, p. viii.

of symbolism expressly designed to facilitate the expression and investigation of truths in the field of chemical composition and reaction. Optimistic philosophers have even hoped to devise algebras that should impart to the treacherous ground of theological and political controversy the solidity and certainty of mathematical inquiry. An invitation from one theological algebraist to another to "sit down and calculate" would then take the place of the denunciations of rival champions inspired with *odium theologicum*.[1]

It is important to note that the aim of all these algebras is the same: namely to correct the weaknesses and supplement the deficiencies of language as an instrument of abstract investigation and exact statement. Words and phrases as the vehicles of ideas are replaced by symbols—with a consequent gain in clearness and conciseness. A formula, consisting of an arrangement of symbols, is free from the ambiguity which often besets the arrangement of verbal units into a sentence, and is, besides, a more effective vehicle of a complicated meaning. It is easier to move forward in an argument when the steps can be expressed in symbolic form, and it is easier to check the correctness of the movements afterwards.

The practical deduction from this general theory of algebra is that in ordinary algebra the symbols are to be thought of as substitutes, not immediately for numbers, but for words as the vehicles of general ideas. The xs, as and bs of a common algebraic statement *refer* to numbers only in the same way as the verbal forms which they replace; they do not *stand* for numbers. To speak of a symbol as a "generalized number" is to employ a phrase which (with all deference to the great algebraists who have used it) has no clear meaning and is incompatible with modern logical ideas. In this book, then, it will be assumed that such symbolisms as $(a + b)^2 = a^2 + 2ab + b^2$, $V = \pi r^2 h$ (for the volume of a cylinder), $y = ax^2 + bx + c$, may always be regarded as verbal statements about numbers expressed for a special purpose in a conventional form, the letters and the graphic symbols being immediately substitutes for words and "standing" for numbers only in the same sense as the verbal units

[1] See the account of Leibniz' Universal Language in Venn's *Empirical Logic*, ch. xxii.

to which they correspond stand for them. This practice will be found to remove much of the difficulty which the beginner has in understanding what algebra is "all about". He cannot easily see how a can mean a particular number without meaning either *this* particular number or *that* one; but he at once appreciates the sport of expressing mathematical statements in a new kind of "shorthand". It is quite true that the miracle of "ambiguous reference" is present in the words "add any two numbers together" in exactly the same way as in the symbolism "$a + b$". But it is so familiar a thing in the former case that nobody but a philosopher finds anything mysterious in it; in the latter case the novelty of the expression brings the strangeness of the fact to our notice.

This point is so important as to be worth repeating in another form. Every one knows that mathematics is essentially concerned with "variables". For instance, in the formula $V = Ah$, which gives the rule for the volume of a cylindrical solid, or in the "equation of the ellipse," $x^2/a^2 + y^2/b^2 = 1$, the symbols all represent variables. That is, while the numerical connexion expressed by the symbolism $V = Ah$ holds good only between particular volumes, areas and heights, the formula refers ambiguously to any set of the volumes, areas and heights which could exist in combination. What is *not* generally noticed is that variables are almost as common outside mathematics as within.[1] Thus in the statement "The King of England is a constitutional monarch" the element "the King of England" is a variable in exactly the same sense as V in the formula $V = Ah$. The sole difference is that while V refers ambiguously to one of an indefinite collection of numbers, "the King of England" refers ambiguously to one of an indefinite collection of persons. "Edward VII" and "George V" are particular values which may be "substituted" for the latter variable, just as the number of cubic centimetres of an actual cylinder may be substituted for the former.

The invention of variables was, perhaps, the most important event in human evolution. The command of their use remains the most significant achievement in the history of the individual human being. Ordinary algebra simply carries to a higher stage of usefulness in a special field the

[1] See Russell, *Principles of Mathematics* (references in index).

device which common language employs over the whole range of discourse. The prudent teacher will, therefore, in the interests of clear understanding and economy of effort, present the technical use of variables in mathematics not as a new thing but as merely a modification of linguistic uses which the pupil mastered, in principle, at his mother's knee.

The use of the sign " = " is sometimes thought to be an obstacle to the view here expounded. How can symbols be anything but numbers, it is argued, when they are connected by a sign of equality? The objection would have more weight if there were not other algebras in which the sign " = " is used to connect symbols which are certainly not numbers: for example, $Mg + H_2O = MgO + H_2$. Consideration shows two common elements in the meaning of the sign " = " in all algebras, namely *identity* and *equivalence*. For example, in the chemical equation just quoted, the sign " = " implies that the "matter" referred to by the symbolism $Mg + H_2O$ and $MgO + H_2$ is identically the same matter manifesting itself in two different forms. The two sides of the equation are therefore equivalent in the sense that one gives us in a different shape all the matter that is contained in the other. Similarly in $(a + b)^2 = a^2 + 2ab + b^2$ or in $y = ax^2 + bx + c$ we have the ideas, first, that the two sides of the equation have reference to identically the same number and, second, that this number can be regarded in two equivalent ways. Thus a chain of symbolical expressions each linked to its predecessor by the sign "=" represents a series of mental occurrences that can be likened to the "transformations of energy" that take place when (say) the explosion of gas in an engine cylinder makes the fly-wheel turn round and so generates in a dynamo an electric current which in turn makes a lamp filament become incandescent. In both cases we have the thought of something which remains fundamentally identical in amount but is made to assume a number of different forms. This is the reason why such a chain of statements as $3 + 5 = 8 + 3 = 11 + 7 = 18 \ldots$ is inadmissible: it implies the introduction of new matter where there ought to be identity of matter throughout.

The practical conclusion from this discussion is that, in ordinary algebra, while the sign " = " always implies the presence of numerical identity beneath changes of form, yet

it may appear in different contexts with varying force. Thus in the chain of statements

$$\frac{d}{dx}(\sin x) = \operatorname*{Lt}_{h \to 0} \frac{\sin (x + h) - \sin x}{h}$$

$$= \operatorname*{Lt}_{h \to 0} \frac{2 \cos \left(x + \frac{h}{2}\right) . \sin \frac{h}{2}}{h}$$

$$= \operatorname*{Lt}_{h \to 0} \cos \left(x + \frac{h}{2}\right) . \frac{\sin \frac{h}{2}}{\frac{h}{2}}$$

$$= \cos x$$

the sign has an obviously different meaning each time it occurs. In the first line it may be regarded as heralding either a definition or a practical rule, in the second it asserts the equivalence of two symbolic expressions, in the third it claims validity for a certain regrouping of the symbols, in the last its force is best given by a simple "is". We may read it "equals" in each case, but it is important, especially in the earlier stages of the subject, to make the pupil realize the varying colour of that convenient word.

The history of algebra throws interesting light on the origin and functions of symbolism. After 230 years it may still be read in the pages of John Wallis as profitably as anywhere. Wallis points out that Vieta (about 1590) greatly improved algebra by first denoting known numbers as well as unknown by "Marks or Notes," and by exercising "all the Operations of Arithmetick in such Notes and Marks as were before exercised in the common Numerical Figures" (*Algebra*, ch. XIV.). Suppose a problem to concern "any three numbers". Then the various numbers in so far as they may be chosen to be the first, the second or the third of the "any three numbers" may be regarded as forming three indefinitely numerous classes. In denoting any member of the first, second or third class by the symbol A, B or C, Vieta followed (says Wallis) the custom of lawyers who "put cases in the name of John an-Oaks and John a-Stiles or John a-Down, and the like, (by which names they mean any person indefinitely, who may be so concerned;) and of later times (for brevity

sake) of J. O. and J. S. or J. D.; (or yet more shortly) of A, B, C, etc." This practice is (like Euclid's use of letters in referring to lines and angles) to save the labour of describing the members of each class "by long periphrases or tedious Descriptions".

In chapter xv. Wallis shows how the English Oughtred (*Clavis Mathematicae*, 1631) "who affected brevity, and to deliver what he taught as briefly as might be, and reduce all to a short view," carried Vieta's "improvement" still farther. "Thus what Vieta would have written

$$\frac{\text{A } \textit{Quadrate}, \text{ into B } \textit{Cube}}{\text{CDE } \textit{Solid}} \textit{ Equal to } \text{FG } \textit{plane},$$

would with him be thus expressed

$$\frac{\text{A}q \text{ B}c}{\text{CDE}} = \text{FG}"$$

From this symbolism to the modern $\frac{A^2B^3}{CDE} = FG$ is but a short though a very important step.[1]

§ 4. *Extended Use of Symbolism.*—A third distinct element appears in algebra when a piece of symbolism originally invented to express a simple arithmetical operation is found so convenient that our definition of the operation is deliberately changed in order to bring other less simple operations within the scope of the same symbolism.

Here again is a feature of such importance that some writers have regarded it as the distinctive characteristic of the science. It is, perhaps, more illuminating to think of it as simply a further development of the power of symbolism to bring efficiency and the "short view". It may be illustrated by continuing a little further our observation of the small boy who was studying areas in § 2. We left him at the point where he had found that the area of a rectangle with integral dimensions is obtained by multiplying the length by the breadth. Let us follow him in his investigation of rectangles whose dimensions are fractional.

There is, to begin with, no difficulty in a rectangle of fractional length but integral breadth. For if the length is, say, $6\frac{1}{4}$

[1] Prof. U. G. Mitchell has contributed a useful review of the growth of algebraic symbolism to Young's *Fundamental Concepts of Algebra and Geometry*, p. 226.

inches and the breadth 4 inches the rectangle can be divided into 4 inch-strips, each containing $6\frac{1}{4}$ square inches. The area will, therefore, be found by taking four times $6\frac{1}{4}$—an operation which comes under the formula "multiply length by breadth". But if the breadth is also fractional—say $4\frac{1}{2}$ inches—the matter is not so simple. The area cannot be divided into an exact number of inch-strips, and cannot, therefore, be calculated by multiplication—as multiplication has hitherto been understood. We are driven to the more complicated calculation:—

$$\text{area} = 6\tfrac{1}{4} \times 4 + \tfrac{1}{2} \text{ of } 6\tfrac{1}{4}$$

Now if the breadth were 5 inches, the calculation of the area would again come under the simple formula "multiply length and breadth together". The formula holds good then, for integral breadths, 3 inches, 4 inches, 5 inches, etc., but not for intermediate breadths, $3\frac{1}{4}$ inches, $4\frac{1}{2}$ inches, $5\frac{5}{8}$ inches, etc. But this is a very unsatisfactory state of affairs. If some of these areas are determined by multiplication it seems reasonable to regard them all as determined by that process. This view is possible if we agree so to extend our notion of multiplication that an operation of the type $6\frac{1}{4} \times 4 + \frac{1}{2}$ of $6\frac{1}{4}$ shall be *defined* as "multiplying $6\frac{1}{4}$ by $4\frac{1}{2}$". By this device the useful formula, *area* = *length* × *breadth*, is made to include every case in which the lengths of the sides of a rectangle can be expressed in integral or fractional numbers.

The introduction of negative and fractional indices is another typical instance of this influence of a useful symbolism upon the original definition of an arithmetical process. Thus, if a sum of money, say £250, increases at 3 per cent compound interest for 5 years its amount is given by the calculation £250 × $(1{\cdot}03)^5$—the symbolism implying (by definition) 5 successive multiplications by 1·03. Similarly, the amount at the end of 6 years is obtained by 6 successive multiplications, symbolized by £250 × $(1{\cdot}03)^6$. Now we must not say that the amount after $5\frac{1}{2}$ years is found by $5\frac{1}{2}$ successive multiplications, for the statement would be meaningless. On the other hand it seems unreasonable that the operation needed in the last case should not be regarded as the same operation and be denoted by the same symbolism as in the other cases. If we yield to this argument we shall agree so to change our original definition of the symbolism

£250 × $(1{\cdot}03)^5$ that the operation (whatever it may be) which determines the amount after 5·5 years shall be capable of being described by the symbolism £250 × $(1{\cdot}03)^{5{\cdot}5}$. In this way all problems of this type can be brought under the single formula

$$\text{Amount} = Pr^n$$

where P is the principal, r the amount of £1 after 1 year and n the time in years, integral or fractional.

The use of negative numbers is, of course, the capital instance in elementary algebra of the characteristic considered in this article.

§ 5. *Manipulation of Symbols.*—Wallis signalized two great merits in Vieta's use of "species" or symbols. First, that they made general arithmetical statements much more concise; second, that the symbols could be "exercised" like numerals, in all the ordinary operations of arithmetic. The second of these properties is the fourth fundamental characteristic of algebra.

It has two chief uses. The first is exemplified by "identities" such as $(a + b)(a - b) = a^2 - b^2$. It is profitable to examine carefully the meaning of this piece of symbolism. It is, we have seen (§ 3), to be regarded as a "shorthand" transcription of a verbal statement: "If any two numbers be selected the product of their sum and difference will be the same as the difference between their squares". It can easily be seen how this result comes about in any specific instance. Consider, for example, the product 15 × 3, expressed as (9 + 6) (9 − 6). Simple arithmetical considerations show that the product can be written in the form

$$\begin{aligned}(9 + 6)(9 - 6) &= 9^2 + 6 \times 9 \\ &\quad - 9 \times 6 - 6^2 \\ &= 9^2 - 6^2\end{aligned}$$

Now this analysis of the way in which the result comes about shows clearly that it is quite independent of the fact that we selected the numbers 9 and 6 for our experiment. The process of multiplication would have taken the same typical course if we had chosen any other numbers. We can, that is, describe the process in words without using the names of any particular numbers, and we can reduce our statement

to a "short view" by expressing it in "species". We then have

$$\begin{aligned}(a + b)(a - b) &= a^2 + ba \\ &\quad - ab - b^2 \\ &= a^2 - b^2\end{aligned}$$

But when we have gone through this process of analysing a particular case and expressing its permanent or essential features by the aid of a symbolic statement we may notice that our work might have been much abbreviated. The symbolism chosen is of such a character that, line by line, it imitates the arrangement of the figures in the arithmetical process. It follows that there was really no necessity to go through the arithmetical process first and afterwards to describe its essential features in symbols. *By manipulating the letters as if they were figures* we could with perfect certainty have predicted the typical form of the arithmetical result.

The "simplification of an algebraic fraction" may be taken as another example of the same property. Suppose that a quantity t is such that $\frac{1}{t} = \frac{1}{4} + \frac{1}{7}$. Then we have

$$\begin{aligned}\frac{1}{t} &= \frac{1}{4} + \frac{1}{7} \\ &= \frac{7 + 4}{4 \times 7}\end{aligned}$$

whence $$t = \frac{4 \times 7}{7 + 4}$$

There are certain features of this process which are obviously independent of the particular numbers which here enter into it. It is possible, therefore, by the employment of "species" to give a generalized account of its steps. Putting a and b for the verbal units "the first denominator," "the second denominator," the three stages of the process can be analysed as follows :—

$$\begin{aligned}\frac{1}{t} &= \frac{1}{a} + \frac{1}{b} \\ &= \frac{b + a}{ab} \\ t &= \frac{ab}{b + a}\end{aligned}$$

But it is obvious that here, as in the case previously examined, the consideration of an arithmetical example is really unnecessary. By treating the letters as if they were figures we might have passed at once and with confidence from the first symbolic statement to the third. That is, we might be sure that if the typical form of an expression for the reciprocal of a number is $1/a + 1/b$, the typical form of the number itself is $ab/(b + a)$.

It should now be easy to understand why "species" are so often taken to be numbers. For on the one hand the symbolic statement about a numerical relation imitates exactly the arrangement of a particular numerical instance of that relation—as, for example, $\frac{1}{a} + \frac{1}{b}$ imitates $\frac{1}{4} + \frac{1}{7}$—and, on the other hand, the general features of the result of transforming a given arithmetical expression can always be predicted by imitating the steps of the transformation with the letters of the corresponding symbolic expression.[1] But while it would be pedantic never to speak of "adding" or "multiplying" symbolic expressions, yet it should be made clear to the pupil that these are merely convenient forms of speech; that, strictly speaking, we can add and multiply nothing except numbers; and that the operations which we carry out upon symbols are merely a means of determining, without considering a specific example, the general or typical features of the result of manipulating an arithmetical expression in a given way.

We may summarize this discussion by saying that from one statement expressed in algebraic symbols it is possible to deduce, merely by manipulating the symbols in accordance with the laws of arithmetic, an indefinite number of statements each of which will be true if the original statement is true. Thus if a number can truly be expressed in the form $(a + b)\ (a - b)$ it is certain that it can be expressed with equal truth in the form $a^2 - b^2$. But by far the most striking instances of this property consist in the transformations which are used in the solutions of equations. Suppose, for example, that I am confronted with the formula

[1] These remarks also make it clear why the Greeks, in the absence of a convenient arithmetical notation, were unable to develop algebra to any considerable extent.

$$C = \frac{nE}{nr + R}$$

taken from an electrician's note-book. Merely by treating the letters as if they were figures I can obtain from this statement a number of other symbolic statements, such as

$$r = E/C - R/n$$
$$R = n\,(E/C - r)$$

Now the important thing here is that I may be so grossly ignorant of electricity that the original formula is meaningless to me. Yet I shall be quite certain that if the original formula was valid the formulæ I have derived from it are equally valid. In this way it is possible for me to discover electrical facts of which my friend the electrician (who though an excellent practical man is, perhaps, but an indifferent algebraist) was actually unaware.[1] In such a way, too, given a symbolic statement of one property of a curve and a few geometrical ideas, it is possible to deduce statements of an endless series of other properties of the curve—properties which may in this way become known for the first time.

Point is given to earlier arguments of the chapter by the observation that the property of yielding new truths as a result of a merely mechanical rearrangement of symbols is not confined to the algebra of number. It is to be found in some form wherever a system of symbolism has been developed "to facilitate reasoning" in a particular province of thought. The late Prof. Jevons actually invented a "logical machine" in which the exploration of the field of truth could be carried out by pulling levers and turning handles. It would probably not be impossible, if only it were worth while, to construct an "algebra machine" which could in a similar way be made to yield from a given formula other formulæ which follow from it.

[1] Of course it will require his practical knowledge to give meaning to my discoveries.

CHAPTER II.

METHOD AND CURRICULUM.

§ 1. *Our Aim in Teaching Mathematics.*—Clear notions about the nature of algebra are essential to a profitable discussion of teaching problems but do not themselves constitute solutions of such problems. The purposes to be held in view in teaching algebra to boys and girls, the proper selection of topics, the best methods of presenting them: these questions constitute a distinct subject of inquiry and are to be settled by considerations of which those adduced in the previous chapter form only a part. For a full discussion of such matters the reader must look elsewhere than in a practical handbook; but he may be asked to accept the following brief statement of general principles as the author's confession of faith.

Mathematical truths always have two sides or aspects. With the one they face and have contact with the world of outer realities lying in time and space. With the other they face and have relations with one another. Thus the fact that equiangular triangles have proportional sides enables me to determine by drawing or by calculation the height of an unscaleable mountain peak twenty miles away. This is the first or outer aspect of that particular mathematical truth. On the other hand I can deduce the truth itself with complete certainty from the assumed properties of congruent triangles. This is its second or inner aspect. The history of mathematics is a tale of ever-widening development on both these sides. From its dim beginnings by the Euphrates and the Nile mathematics has been on the one hand a means by which man has constantly increased his understanding of his environment and his power of manipulating it, and on the other hand a body of pure ideas, slowly growing and consolidating into a noble rational structure. Progress has brought about, and, indeed, has required, division of labour.

A Lagrange or a Clerk Maxwell is chiefly concerned to enlarge the outer dominion of mathematics over matter; a Gauss or a Cantor seeks rather to perfect and extend the inner realm of order among mathematical ideas themselves. But these different currents of progress must not be thought of as independent streams. One never has existed and probably never will exist apart from the other. The view that they represent wholly distinct forms of intellectual activity is partial, unhistorical, and unphilosophical. A more serious charge against it is that it has produced an infinite amount of harm in the teaching of mathematics.

Our purpose in teaching mathematics in school should be to enable the pupil to realize, at least in an elementary way, this two-fold significance of mathematical progress. A person, to be really "educated," should have been taught the importance of mathematics as an instrument of material conquests and of social organization, and should be able to appreciate the value and significance of an ordered system of mathematical ideas. There is no need to add that mathematical instruction should also aim at "disciplining his mind" or giving him "mental training". So far as the ideals intended by these phrases are sound they are comprehended in the wider purpose already stated. Nor should we add a clause to safeguard the interests of those who are to enter the mathematical professions. The treatment of the subject prescribed by our principle is precisely the one which best supplies their special needs.

§ 2. *The Choice of Curriculum.*—This principle adopted, we have the practical problem of selecting for our curriculum in algebra topics which shall illustrate adequately the two aspects of mathematics. To the question whether the two must be provided for separately we may reply with a confident negative. The theoretical questions which are of most importance in an elementary course are just those which arise naturally out of attempts to apply mathematical ideas and methods to practical purposes. We shall, therefore, choose those subjects in which the practical value of algebra is most clearly exhibited, confident that this is the best means of securing opportunities for fruitful theoretical discussions.

In accordance with this view algebra should be introduced to the young pupil as a symbolic language specially adapted for making concise statements of a numerical kind about

matters with which he is already more or less familiar. In other words, the earliest lessons in the subject should teach the use of the formula, illustrations being drawn largely from the "Pocket Book" of the engineer and similar formularies. In these lessons the various algebraic notations should be introduced as symbolic idioms needed for the transcription of pieces of important or interesting fact. A little later comes the study of those manipulations of a formula by which it may be made to yield truths unknown or unperceived before (ch. I., § 5). It is of the first importance that from the outset of his work the pupil should be made to perceive clearly and feel constantly that both formulæ and manipulations always refer to realities beyond themselves. It is scarcely too much to say that incompetence in, and distaste for, mathematics nearly always spring from the neglect of this fundamental teaching principle. Even in the case of those who have a natural fondness for the technique of mathematics the same neglect often leads to an astonishing blindness to the real significance of mathematical ideas and operations.

The gradual elaboration of the formula as an instrument of description and investigation is, then, the first business of the course in algebra. The pursuit of this task leads naturally to two kinds of inquiries both of which belong to the inner aspect of mathematical thought. In the first place, the attempt to build up an effective system of symbolic expression leads to a number of problems of great importance from the theoretical point of view. These necessarily include questions about the meaning and use of negative numbers, the interpretation of fractional indices, etc., and may range as far as the fascinating inquiries into the ultimate nature of numbers and numerical operations which are so characteristic of modern mathematics. In the second place, in trying to give an account from the numerical standpoint of the concrete things with which his formulæ deal, the young algebraist can hardly fail to notice and to become interested in the fact that "variables" of widely differing character are yet often bound to one another by identical quantitative laws. From that moment onwards it is natural to give an increasing amount of attention to these general forms of connexion between variables. Eventually—under the rather forbidding name of "functions"—they may become the main object of study.

§ 3. *The Position of Trigonometry and the Calculus.*—The foregoing principles point directly to two important modifications in present teaching practice. The first concerns the position of trigonometry, the second that of the differential and integral calculus. At present it is usual to teach both of these as "subjects," distinct from algebra and distinct from one another. The reason is easily found. The professional mathematician, interested chiefly in perfecting the technique of his subject, finds it natural as well as most effective to take a special group of allied methods or allied problems and to develop them as far as he can without concerning himself too greatly about the practical value of his work. Now this systematic exploration of special parts of mathematics is, no doubt, of vital importance for the continued growth of the science. It does not, however, by any means follow that the branches of mathematics should be presented to beginners with the formal elaboration which is the inevitable mark of their treatment as separate subjects. The absurdity to which this practice may lead is sufficiently illustrated by the custom of one very important public examination in which, until recently, a paper was set requiring considerable technical knowledge of the differential calculus while assuming no knowledge at all of integration. The mischief which follows from it is still more evident from the fact—surely little less than a scandal—that so far as the operation of University Leaving Certificates is concerned a boy or a girl may at present [1913] pass through a secondary school without making the smallest acquaintance with the fascinating and powerful methods of elementary trigonometry.

If we admit that the custom of teaching trigonometry and the calculus as distinct subjects has no defensible foundation and is responsible for very unsatisfactory consequences, we shall be prepared to inquire what position is assigned to them by the principles laid down in this chapter. Taking first the case of trigonometry, it becomes at once evident that the pupil's acquaintance with the tangent, sine and cosine should begin in the region where arithmetic marches with elementary geometry. As soon as the symbols of the trigonometrical ratios are recognized as capable of entering into formulæ and of being manipulated they should be regarded as belonging to the vocabulary of algebra. There is, indeed, no principle, except the invalid principle of formal segrega-

tion, upon which we can include the study of x^n or of a^x in the algebra course and exclude $\sin x$ or $\tan x$. All alike are pieces of symbolism invented for the description and interpretation of facts of the external world. Each represents a typical kind of "function". To each corresponds a specific form of curve which may be regarded as the graphic symbol of the function. Both algebra and trigonometry would gain by fusion: the former through an added variety and richness in the illustrations of its main themes; the latter by the removal of the excessive formalism which at present obscures its value and interest for the beginner. Fusion upon these lines is attempted in this book. The formal work which constitutes so large a part of the ordinary treatment of trigonometry is greatly reduced and much more stress is laid in Part I upon the practical, and in Part II upon the functional, aspect of the subject. Among other advantages this plan permits the inclusion of a brief section upon the trigonometry of the spherical surface. The customary neglect of a branch of inquiry which is of such importance and interest, in which also the essential results are to be obtained with such ease, is, of course, only another glaring instance of the mischief brought about by the over-formal treatment of mathematics in school. There is absolutely no reason why the main elements of the trigonometry of the sphere should not be taught except the ridiculous one that spherical trigonometry is a separate "subject" which can be studied only after "plane trigonometry" has been disposed of.

In arguing for the assimilation of trigonometry in the algebra course we are arguing against the artificial separation between problems which must in any case be attacked in a similar spirit and by similar weapons. When we consider the position of the differential and integral calculus we have to protest against a tradition which forbids all but exceptional pupils to become acquainted with the most powerful and attractive of mathematical methods. As in the former case, the mischief is the result of a technical elaboration which, though essential to the historical development of the calculus, has had the effect of making the really simple ideas upon which it is built inaccessible to the ordinary boy or girl at school. In this instance the history of the subject suggests a remedy for a state of things which is generally regarded

as unsatisfactory. The calculus began, in the writings of John Wallis and others, merely as a special kind of algebraic argument which might be introduced at any appropriate point and without the apparatus of a technical notation. The remedy suggested by this observation is adopted in the present work. Notions which form part of the doctrine of the calculus are introduced at an early stage and are developed side by side with other algebraic ideas ; but only towards the end of the work are the technical symbols introduced which have been known to so many students only as hostile standards floating above an impregnable citadel.

§ 4. *Some Practical Suggestions.*—It is probable that enough has been said to indicate the point of view adopted in this book, and there would be nothing to gain by anticipating here what will be found in detail in later chapters. The whole course treated in these chapters is divided into two main stages upon a principle already explained in the Preface. These stages are represented respectively by Parts I and II of the *Exercises in Algebra*. The ground assigned to the two stages, their subdivisions, and the order of treatment recommended are dealt with in ch. v. as far as concerns Part I, and in ch. xxxviii. as far as concerns Part II. The reader who wishes to gain a complete view of the whole course may with advantage turn immediately from this chapter to those.

Meanwhile it may be convenient to give some account of the plan of the book and to suggest how it should be used. The present or Introductory Section contains two more chapters. These deal in some detail with the two instruments of expression which the student of algebra has to use at every stage of his progress—namely, the formula and the graph. The conclusions reached are applicable all through the course, but they have special reference to the first beginnings of algebra. Thus they are illustrated largely by references to Exercises I and II, which are intended to be regarded as preliminary work to be covered before the systematic study of the subject is begun.

The rest of the book—setting aside the general introductions to Part I and Part II—is divided into sections which correspond to the ten sections of the exercises. In Part I each of these divisions contains a chapter or chapters in which the general ideas underlying the exposition of the section are explained, and hints are given for the illustration

of the teacher's lessons, etc. These chapters also offer remarks upon the examples of each exercise. Matters which require emphasis are pointed out and the solutions recommended for typical examples are given. The remaining chapters indicate the substance of the lessons which the teacher should give before setting his class to work upon the various exercises. He should understand that these notes do not always contain a complete discussion of the topic with which they deal. They aim merely at carrying the discussion up to the point at which the pupil can profitably attack the exercise. In many cases the subject is developed farther in the course of the exercise itself. For this reason alone it is important that the teacher should take occasion, after an exercise has been disposed of, to summarize afresh what has been learnt both from the preliminary lesson and from the subsequent examples. It may be added that sometimes the examples, instead of developing farther a topic which has received preliminary discussion in class, prepare the way for a subsequent lesson. In such cases warning is always given that certain examples are specially important, and what is to be learnt from them is summarized in the chapter devoted to the following lesson.

Part II is written upon a similar plan, with the important difference that there are no systematic suggestions for lessons preliminary to the exercises. These are unnecessary in view of the detailed discussions which accompany the exercises.

With regard to the use of the exercises many points must be left to the teacher's own discretion and experience. For instance, it is hardly ever desirable that the whole class shall at the first attack work through all the examples of an exercise. Certain examples must be taken because, as we have said, the theory of the subject is developed from them, while others are essential as a preparation for instruction to come later. The teacher must use his judgment as to the number of examples, outside these, to be exacted from every pupil. Some may be left for subsequent homework, some for revision in class, some (especially in the harder sections of the exercises) for the cleverer and quicker boys or girls. The teacher should bear in mind that it is often a good thing to run away from a difficulty in order to fight it another day; that of many a difficulty, especially in mathematics, it may be said, *solvitur ambulando;* and that a class

bored by long-continued study of a single topic is making its minimum rate of progress and gaining the minimum profit from its labours.

To these remarks the author may, perhaps, be permitted to add for the benefit of his younger colleagues that it pays *always* to exact from a boy or girl the best work of which he or she is capable. The beginner, recognizing that his real business is to keep his pupil's mind active, is apt to underestimate the closeness of the connexion between orderly ways of thinking and writing, and to be imposed upon by specious little rogues who have developed a precocious talent for concealing laziness. He should be on his guard against this weakness, and should constantly check any tendency to accept careless or untidy work merely because it shows intelligence. When he has acquired the diagnostic powers which only experience can give he may trust his ability to determine the cases in which he may safely relax. Even then these cases should be rare.

References for Reading.

The present author has dealt more fully with some of the points raised in chs. I. and II. in the following papers :—

"On the Method of School Algebra." "School," Sept. 1905. (John Murray, 6d.)
"The Arithmetic of Infinites." "Mathematical Gazette," Dec. 1910 and Jan. 1911. (Bell & Co., 1s. 6d. each.)
"The Aim and Methods of School Algebra." "Mathematical Gazette," Dec. 1911 and Jan. 1912.

The following are among the most important and accessible books upon the logic, pedagogy and history of mathematics. Some of them contain bibliographies :—

A. N. Whitehead, "Introduction to Mathematics". (Home University Library, Williams & Norgate, 1s.)
P. E. B. Jourdain, "The Nature of Mathematics". (The People's Books, T. C. & E. C. Jack, 6d.)
J. W. A. Young, "Lectures on Fundamental Concepts of Algebra and Geometry". (The Macmillan Co., 7s.) An admirable review of the subject, intended specifically for teachers.
Bertrand Russell, "The Principles of Mathematics". (Cambridge Univ. Press, 25s.) A book of the highest originality, importance and authority.

B. Branford, "A Study of Mathematical Education". (Clarendon Press, 4s. 6d.) The most important and original of recent English contributions to the pedagogy of mathematics.
D. E. Smith, "Teaching of Elementary Mathematics". (Macmillan Co., 4s. 6d. net.)
J. W. A. Young, "Teaching of Mathematics in the Elementary and Secondary School". (Longmans, 6s.) The last two are excellent American textbooks.
G. St. L. Carson, "Essays on Mathematical Education". (Ginn & Co., 3s. net.)
J. Perry, "Report of a Discussion on the Teaching of Mathematics," British Association, 1901. (Macmillan & Co., 1s.) A Report which has had a great influence upon the reform of mathematical teaching in England.
A. Höfler, "Didaktik und Methodik des Rechnen und der Mathematik". (Leipzig, Teubner, 12s.)
W. W. Rouse Ball, "Short Account of the History of Mathematics". (Macmillan & Co., 10s.)
F. Cajori, "A History of Elementary Mathematics". (The Macmillan Co., 6s. 6d. net.)
J. Tropfke, "Geschichte der Elementar-Mathematik," 2 vols. (Leipzig, Veit & Co., 9s. each.)

CHAPTER III.

THE FORMULA.

§ 1. *Generalized Arithmetic.*—Algebra regarded as "generalized arithmetic" should have no formal beginning. As soon as the child who sees the teacher write upon the blackboard

$$\text{area} = \text{length} \times \text{breadth}$$

can translate this into the words: "To find the area of the floor I must multiply its length by its breadth,"[1] he has, without knowing it, already begun his study of the subject. What the teacher has set before him has the two characteristics of a "formula": (*a*) it is a statement of a general "rule" applicable to any one of a definite class of problems; and (*b*) the statement is expressed in a conventional form chosen for its properties of conciseness and ready comprehensibility (ch. I., §§ 2, 3). By his twelfth year lessons in arithmetic and science should have afforded the pupil abundant opportunity of learning to write down and use simple formulæ of this kind.

Formulæ must, of course, never be used in arithmetic unless the pupil clearly understands the processes which they prescribe. In other words he is entitled to use a formula only if it represents genuine results of his own thinking. He may then with advantage write it at the head of his calculation as a memorandum of the process which he intends to employ. Used in this way, the formula makes for greater clearness both of the pupil's thinking and of his written statements.

These preliminary exercises will give occasion for the use

[1] The purist may make him say: "the number which measures its length by the number which measures its breadth". This greater scrupulosity of diction need not affect the formula.

in verbal formulæ of all the ordinary symbols of arithmetic: e.g. $\frac{\text{area}}{\text{length}}$, or area/length, (length)^2, (length)^3. The teacher must also seek in them his opportunity for introducing his pupils to the extremely important device of replacing the constituent phrases of the verbal formula by single letters. The use of words in an abbreviated form supplies a natural transition to the stage in which this practice is definitely adopted.

§ 2. *Use of Literal Symbols.*—The use of single letters can best be explained (and made attractive) by teaching the class to regard formulæ as "shorthand" memoranda of the rules which they have established in the course of their work and are constantly needing. The principles of this "shorthand" are (*a*) to represent certain constantly recurring words (such as "multiply," "divide," "square") by conventional symbols or "grammalogues"; and (*b*) to reduce other words or verbal expressions in the full statement of the rule to single letters, chosen so as to suggest those words or expressions as readily as possible to the reader of the memorandum. Thus "circumference of the circle" may be reduced to C, "rate of interest per cent per annum," to r, "the number of passengers" to either n or p, according to convenience or the choice of the writer. Moreover, it will be in accordance with the notion that we are developing a "shorthand" to replace always by the same letters any words or verbal expressions which frequently occur in the problems in which our formulæ are employed. Thus A can generally be taken without special explanation to mean "area," the particular figure whose area is in question being known from the context. This method of procedure follows so obviously from the function of the formula as a labour-saving device, that it is, perhaps, unnecessary to point out how it accords with the practice of all persons who employ formulæ for serious purposes. To the electrical engineer or the actuary it is a matter of no small importance that his text-books and formularies should employ consistently the same symbols, and that these should readily suggest the verbal units for which they stand.[1]

[1] The Institute of Actuaries have published an official set of symbols which are used, without explanation, in all the papers and discussions of their members.

When the verbal statement of the rule contains a numerical constant the practice must be taught of placing it before the literal symbols. Thus the rule that the volume of a pyramidal solid is obtained by multiplying the area of the base by one-third of the height is, in accordance with this convention, to be written neither in the form $V = Ah\frac{1}{3}$, nor in the form $V = Ah\frac{1}{3}$, but in the form $V = \frac{1}{3}Ah$.

The use of the symbol π appears to contradict the statement that symbols are to be taught as representing not numbers but words or verbal expressions. This is not really the case. There would be no point in using the symbol π in a formula, rather than a concrete number, if it was not understood by the class that the ratio signified is one that in different calculations is taken to have different values according to the degree of approximation required. Thus π does not immediately represent a number but is the "shorthand" rendering of the phrase: "the ratio of the circumference of a circle to its diameter, taken to the degree of approximation which the problem requires".

§ 3. *Example of Method.*—It is a matter of little importance at what precise point the class should be taught to adopt definitely the device of representing by a single letter a word or larger verbal unit. Whenever the step is taken the teacher will find it convenient to proceed much as in the following example.

The topic under discussion is supposed to be the mode of calculating the volume of a solid of uniform cross-section, e.g. a cylinder. The teacher is provided with a cylindrical tin over the bottom of which he has pasted a piece of paper divided into centimetre squares. He proposes to the class the problem of finding the number of cubic centimetres in the content of the tin and conducts with them a colloquy to the following effect:—

On counting up the number of square centimetres that cover the base of the tin (making due allowance for incomplete squares) I find that there are (we will say) exactly 32. If now I placed in the tin a layer of clay 1 cm. thick and just large enough to cover the base I should evidently have a centimetre cube of clay standing on each of the squares. There would be, therefore, 32 c.cm. in the layer altogether. If the tin is (say) 15 cm. high, I can pack 15 of such layers of clay on top of one another, so that the total amount of clay

that the tin would hold must be 32 × 15 c.cm. Now if there had been on the base of the tin 17 or 148 or any other number of square centimetres and 82 or 2003 or any other number of linear centimetres in its height it is evident that I could calculate the amount of clay the tin would hold in exactly the same way. Thus we have the rule that the number of cubic centimetres which the tin would hold (the "volume" of the tin) is found by multiplying the number of square centimetres in the area of the base by the number of linear centimetres in the height—or, expressed more shortly, by multiplying the area of the base by the height. If the area of the base had been fractional—say 32·7 sq. cm.—each of the slabs of clay would have contained 32·7 c.cm. instead of 32 c.cm. If, moreover, the height of the tin had been 15·3 cm., instead of exactly 15 cm., then it is clear that I could have packed into the tin 15 slabs and a thinner slice of clay 3/10 of a centimetre thick, and therefore containing 3/10 of 32·7 c.cm. Thus the total volume of the clay would be given by the product 32·7 × 15·3. The rule that the volume of the tin is obtained by multiplying the area of the base by the height evidently holds good, then, when the area or the height is measured by a fractional number as well as when the measures are whole numbers.

We may now proceed to write this rule upon the blackboard. It is unnecessary to write every word in full, for you will have no difficulty in knowing what I mean if I shorten it down to the following:—

$$\text{vol. of cyl.} = \text{base} \times \text{height}$$

or if I make it briefer still:—

$$\text{vol. of cyl.} = \text{base} \times \text{ht.}$$

Now there are a great many persons who have constantly to make use of notes or memoranda of this kind. They are such people as engineers, who have to keep notes of all sorts of rules in regard to the weights which their materials will bear, etc., ship-builders, electricians, architects, military officers, sailors, etc., etc. Some of their rules are so complicated that their notes would be very cumbersome even if they shortened the words down as we have, and employed symbols such as "=" and "×". They find it necessary to use, therefore, a kind of shorthand in which they can express their memoranda much more briefly even than we have expressed

the rule for finding the volume of the cylinder. The principle is to use one letter only to represent a word such as "height" or a phrase such as "area of the base". As far as possible letters are chosen which readily suggest the words for which the letters stand. Thus if we were to write our formula in this "shorthand" way we could choose the letter V to stand for the words "volume of a cylinder," the letter B to stand for the words "area of the base" and the letter h to stand for the word "height". The rule would then take the extremely short form :—

$$V = B \times h$$

But if we make up our minds never to use more than one letter to represent a word or group of words, the formula may be shorter still. We can agree to indicate that two numbers are to be multiplied together simply by writing the letters which are the shorthand descriptions of them side by side. Upon this plan our formula becomes :—

$$V = Bh$$

Such a formula is, remember, merely a shorthand way of writing down the sentence "The volume of the cylinder is obtained by multiplying the area of the base by the height". The sign "=" can be read "is obtained by" (or equivalent words), while the word "multiply" is supplied by the fact that the letters are side by side. If it is necessary to make clear in your note-book what words the various letters stand for it is best to write as follows :—

$$V \equiv \text{"volume of a cylinder"}$$

Thus the symbol "$\equiv$" is to be read "is the symbol for the words".

From this point onwards the teacher should take occasion whenever a rule in connexion with mensuration, arithmetic, or elementary science has been formulated in words to discuss with the class how to express it in "shorthand" form. In this way the class will, without receiving special lessons in algebra, acquire facility in handling the simpler forms of symbolical expression.

Exercise I is intended to indicate the range of algebraic expression which should be covered in these incidental discussions. It will, therefore, be most conveniently used as a

means of revising and assuring the pupil's preliminary knowledge before he begins, in Ex. III, the systematic study of the art of formulation. If the pupil has no knowledge of the uses of the formula, Ex. I must be worked through with considerable care. In either case it is of much importance that the method recommended in this chapter should be followed; that is, that the beginner should be taught to regard a formula as nothing more than a "shorthand" transcription of a verbal rule or other statement.

§ 4. *Setting Down of Work.*—By writing his formulæ at the head of arithmetical calculations the young student will also learn the rudiments of the art of "substitution". In connexion with this topic it is hardly possible to lay too much stress upon the importance of cultivating a neat and orderly way of setting down the steps in an arithmetical or algebraical argument. A piece of algebraic symbolism should be as capable of straightforward and continuous reading as a passage from a newspaper. To achieve this end the teacher will find it a sound rule never to permit a line to contain more than two expressions connected by the sign of equality, and to insist upon the pupil's setting the signs of equality, in successive lines of the argument, directly underneath one another. Thus such expressions as

$$V = Bh = 32{\cdot}7 \times 12{\cdot}4 = 405{\cdot}48 \text{ c.cm.}$$

should be excluded both from the exercise book and the blackboard in favour of the arrangement:—[1]

$$\begin{aligned} V &= Bh \\ &= 32{\cdot}7 \times 12{\cdot}4 \\ &= 405{\cdot}48 \text{ c.cm.} \end{aligned}$$

§ 5. *No Manipulation of Symbols.*—It should be noted that in this preliminary work no manipulations of the symbols will be taught and no question of the "sign" of the numbers that are substituted in the formulæ will be raised. Both these matters belong distinctly to the formal study of algebra.

[1] The need of economizing space unfortunately compels the author to break this rule in his book. The reader should attend to the precept and ignore the examples.

CHAPTER IV.

THE GRAPH.

§ 1. *The Graph and the Formula.*—Throughout algebra the graph, alone or in conjunction with the formula, plays an important part as an instrument of analysis and generalization. It is obvious that it shares many of the properties of the formula. Like the formula it can be used to bring out and express the "law" or identity which underlies the diversity of a number of concrete numerical facts. Like the formula it delivers its message in a form readily taken in by the eye, and so ministers to the "short view". It may often be regarded as a general statement from which, as by substitution in a formula, an endless number of new particulars can be deduced. Lastly, it can in some cases be manipulated like a formula so as to yield new and unsuspected generalizations. On the other hand, it is inferior to the symbolic formula in many important respects. Its accuracy depends largely upon mechanical or non-intellectual conditions, such as the skill of the draughtsman and the exactness of the squared paper. It is less compact and less easily reproduced. Its message is frequently inarticulate and obscure. For these and similar reasons it should be regarded as a *subsidiary* algebraic instrument which fulfils its best office when it either leads up to a formula by which it may itself be superseded, or serves to unfold more fully the implications of a formula whose properties have been only partially explored.

As contrasted with the symbolic formula the usefulness and the limitations of the graph both rest upon the same circumstance—its relatively concrete character. Thus the curve which represents the relations between the values of two variables is abstract enough to be a means of concentrating attention upon the law of connexion as distinguished from the variables connected, and is at the same time concrete

enough to make a vivid appeal to intuition. On the other hand, though its concreteness makes it a more impressive form of statement than the formula, it also makes it a much less flexible instrument of investigation.

In view of these characteristics it is not surprising to find that graphic methods were used as an effective instrument of mathematical thinking before algebraic symbolism had developed beyond the rudiments. Thus the Greeks, who never succeeded in producing a satisfactory algebraic method, yet performed analytic feats of high importance with the aid of graphic forms. The propositions of Euclid's second, fifth and most of the later books exemplify this statement; a simpler, but very striking illustration, is afforded by the argument in proof of the rule for extracting a square root which is given in ch. VIII. This method (as the present author found after publishing it as his own!) was used, in practically the same form, by the Alexandrian astronomer, Ptolemy, about A.D. 120.[1]

The superior vividness and intuitability of graphic modes of expression suggests the conclusion that the young pupil should be taught their simpler uses before he makes acquaintance with the abstruser though more powerful instrument, the formula. The historical circumstances just adduced will seem to many to add support to this conclusion. It is true that it is opposed to the prevailing practice, but no one is likely to contest it who has observed the readiness with which a child will express himself in pictures long before he has learnt to command the more abstract medium of written words. Thus the gradual penetration of graphic methods into elementary instruction in mathematics and science may be welcomed as one of the most significant features of present pedagogical tendencies.

§ 2. *First Lessons in Graphic Representation* (Ex. II, A.). —In accordance with the observations of the last article the second of the two preliminary exercises (Ex. II) consists of a number of examples illustrating those uses of the graph which foreshadow the more effective use of the formula to be learned at a later stage. Like those of Ex. I these examples are best used as a means of revising and extending a know-

[1] See the *School World*, Feb. 1911; also Heath, *Archimedes*, p. lxxvi.

ledge of graphic representation which the pupil acquired before he began the systematic study of algebra. They summarize, therefore, a course of instruction which will be all the more useful if it has been spread over two or three years of school life. If the pupil has not been practised in graphic representation in his earlier lessons in arithmetic and geometry this exercise will require special attention. It is, however, neither necessary nor desirable to work through all the examples before going on to Ex. III. A few of the examples in divisions A and B should be worked in class. The rest can be set from week to week as homework, or worked in class alternately with the first few examples of Ex. III. The only thing essential is that the main principles of graphic representation (i.e. those exemplified in divisions A and B) should be well understood before the class attacks Ex. IV.

The best way to teach the graph is to let it grow out of the use of the picture. This principle is illustrated in the examples of Ex. II. Thus division A begins with graphs which are very little removed from pictures, and goes on to others in which the pictorial element is constantly less prominent, until, in the examples of division B, the pupil is prepared to dispense with it almost entirely.

Nos. 1, 2 and 4 are exercises which form part of a course of lessons in Nature study given to children of about nine years old. (The author owes them to the kindness of his colleague, Miss C. von Wyss.) In **No. 1** a base line is divided into equal parts and graduated to represent minutes. In the middle of each interval a perpendicular line is drawn of the same length as the path traced out by the snail during the corresponding minute. This length is to be determined by laying a piece of cotton thread along the line in the diagram. Fig. 1 [1] shows the solution of **No. 2.** Vertical lines are drawn through the minute graduations of the base, and the perpendiculars drawn in No. 1 are placed end to end across the spaces between them. The firm sloping lines represent in this way the movements of the snail which traced the path AB ; the broken lines represent those of his competitor. The variations in the average speed of the snails are, of course,

[1] I.e. in this book. Figures in *Exercises in Algebra*, Part I and Part II, are numbered consecutively among themselves. A reference to one of them will in this book always be prefaced by "*Exercises*" for Part I, and by "*Exercises II*" for Part II.

expressed by the variations in the slope of the lines. Thus the pupil acquires in his first graphical exercise the germ of the idea of measuring a rate of change by the "gradient" of a graph. On the whole, however, these two examples illustrate the lowest grade of usefulness of a graph. The diagrams add practically nothing to the information contained in the original figures; they merely present it in a more effective and easily assimilable shape.

FIG. 1.

From this point of view **No. 4** represents an important step forward, for it introduces the process of *interpolation*. Having graduated his base-line as before, the pupil sets up at the end of each segment a vertical equal to the length of the tulip as shown in the corresponding drawing in *Exercises*, fig. 2. A vacant place must, of course, be left at the end of the third segment. The pupil is then to judge how long the tulip would have been found to be if this particular measurement had not been omitted. *After* (rather than before) he has expressed his judgment the curve through the tops of the ordinates is to be drawn, and it is to be made clear that, consciously or unconsciously, the course of this curve really determines the judgment.

The significance of the process of interpolation consists in the draughtsman's analysis, out of the given lengths of the tulip, of the law of succession which they suggest. He may be quite unable to give a precise account of this law—he may even not understand what is meant by calling it a law—but the graphic presentation of the data forces it upon his attention as an actual fact. **Nos. 5** and **6**, taken together, are meant to fortify the incipient notion of a law by contrasting a case where law exists with one from which it is absent. In No. 5 the "lawfulness" in the growth of the tulip leads the pupil to expect a corresponding "lawfulness" in its weekly increase. We may permit him at this stage to act upon the expectation, even though he cannot give an adequate defence of it. Thus when the growth during the fourth week is determined by interpolation he will expect it to be the same

as the difference between the length of the plant at the end of the third week and the length inserted by interpolation in No. 4. On the other hand it is evident that there is no law in No. 3 and, therefore, that nothing can be deduced by interpolation.

An important technical detail must not be forgotten. The vertical [1] lines in **No. 4** represent the height of the plant at the end of each week; they should be drawn, therefore, at the end of the corresponding segments of the base line. On the other hand the verticals in **No. 5** represent growth *during* the interval and should, accordingly, be erected at the mid-points of the time-intervals.

Of the remaining examples in division A nothing need be said except that **No. 7** describes an elementary type of astronomical observations which may with great advantage be carried out practically. Such observations lead to the recognition (i) that the sun reaches its highest point in the sky at a time which varies from about a quarter to twelve to about a quarter past by local time, and (ii) that when at its highest point it is always exactly in the same direction—the direction called "south".

§ 3. *Ex.* II, *B.*—In division B the pictorial element present in all the examples of division A disappears. Instead of thinking chiefly of his representations of the data the pupil is now to attend directly to the graphic expression of the law which governs their variation. The "note" before **No. 11** is intended to prepare him for this shifting of the centre of interest. The teacher will see that there is an important gradation in the abstractness of the examples. In the earlier ones the ordinates, if they were drawn, would represent the data pictorially, for the data are all lengths. In the later ones the ordinates do not represent lengths, and, therefore, cease to be even implicitly pictorial. Much of the difficulty which beginners experience in understanding the representative character of a graph is met with at this point. The teacher does not always make allowance for the gap which the child feels between the case in which a length represents

[1] The terms "horizontal" and "vertical" will be used throughout this book to denote the directions upon a sheet of paper in which the axes of x and y are respectively drawn according to the common usage. They may be taken to refer originally to lines drawn upon a blackboard arranged with its plane vertical.

a length and the case in which it represents something totally different from itself. The note prefacing **No. 15** suggests a way of bridging the gap. In this example the ordinates are to represent sums of money, and the pupil is told to think of each of them as showing the height of a certain pile of shillings or pence. Such devices facilitate the transition to the stage in which the abscissæ and ordinates become purely representative and the graph an expression of an entirely abstract law of connexion between the numerical values of variables.

When a graph ceases to be pictorial we need a principle to decide which of the two variables shall be represented by horizontal measurements and which by vertical. There is a perfectly clear rule upon this point which the pupil should be taught to apply unaided. In every case in which a graph is drawn we can regard one set of measurements as having been *chosen*, and the question is how the other set of measurements depends upon these. Thus in **No. 15** we select from the dealer's catalogue statements about the monthly payment demanded for furniture of a certain total value, and the question is how does the former sum depend upon the latter. The former may be called the *dependent* variable, the latter by contrast the *independent* variable. Then the rule in question states that the independent variable is to be represented by horizontal, and the dependent variable by vertical measurements. The teacher may choose to add the information that the former measurements are called *abscissæ* and the latter *ordinates*, but these technical terms are probably better reserved until a later point.

§ 4. *The Column-graph* (Ex. II, C.).—So far the variables whose connexion is the object of inquiry have been in each case represented in the graph by measured lengths. There have been, however, certain cases in which this mode of representation is not completely satisfactory. No. 5 offers a typical instance. Here the thing to be represented is certainly a length, but it is a length acquired by the growing tulip, not at any particular moment of time, but gradually during the course of the week. There is therefore something arbitrary in connecting the representative line with the middle or any other particular point of the base-segment. It is obvious that a more satisfactory mode of representation would connect the weekly growth with the whole of the segment representing the time in which it accrued.

The best way to meet this objection is to represent each weekly increase by an *area* erected upon the corresponding segment of the base. But the representation of a length by an area is more difficult to understand even than the representation of a sum of money by a length. It is important, therefore, to introduce this new form of graphic symbolism by examples in which it is readily intelligible. This condition is fulfilled by **Nos. 21** and **22**, the solution of which is shown in fig. 2. The horizontal axis is here divided into equal segments to represent intervals of one day. Upon each segment is set a rectangle whose area measures the amount of fresh ground swallowed up by the flood during the corresponding day. A larger square of the paper is taken to represent an acre. When completed the series of rectangles constitutes a "histograph" or (as we shall prefer to call it) a "column-graph" representing pictorially what we are told about the history of the flood. It is obvious that the total area flooded at the end of the first, second, third, . . . days is represented by the combined area of the first, the first two, the first three, . . . rectangles.

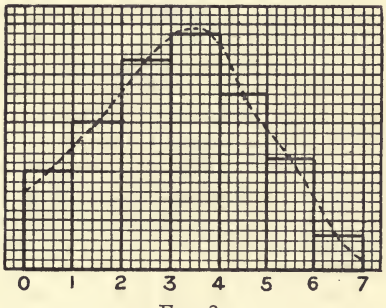


FIG. 2.

The data must in the first instance be represented by rectangles, for the reason that they consist merely of statements of the total increase of the area of the flood during each day. We are supplied with no direct information about the way in which the flood grew from hour to hour. Yet as contemplation of the pictures given in *Exercises*, fig. 2, inevitably suggests dimensions for the missing member of the series, so contemplation of the gross results of the successive days' flooding inevitably suggests probable details of its history. We may assume, to begin with, that the growth of the flood was continuous, and we *know* how much was added to its surface during the course of each day. The assumption com-

bined with the knowledge suffice to suggest the interpolations represented by the broken curve which is added in fig. 2 to the original column-graph. The principle followed is to draw a continuous curve across the rectangles in such a way that the total area above each of the original segments of the base remains unchanged. It is possible, of course, to draw many curves which fulfil this condition, but if we qualify it by the further condition that the curve shall be as smooth as possible —that is, if we assume the changes in the rate of flooding to have been as little violent as the data permit us to suppose— the results obtained by different draughtsmen will show a satisfactory agreement. To answer such a question as **No. 22** (i) we shall, of course, refer to the column-graph as corrected by the added curve. The total area under water at the end of $1\frac{1}{2}$ days is represented with much probability by the area under the curve from the ordinate at 0 to the ordinate at 1·5.

The column-graph is the means of representation which should be used whenever the data state, not the values of one variable corresponding to definite values of the other, but the gross amount of the change in the first variable corresponding to a given change in the second. As in the case of the linear graph, the representation varies with the nature of the variables from a pictorial to a purely symbolic stage. The earlier stages in the gradation are illustrated by the examples of division C. It is unnecessary to delay the progress of the class in order to work through these examples. It will be sufficient, and probably best, to take them from time to time side by side with Exs. III-XV. They will be found of great importance in later stages of the course, for they are not only constantly used by modern statisticians but are also made in this book to play an essential part in the exposition of the main ideas of the differential and integral calculus. Thus the method is one which the pupil will often be required to use.

§ 5. *Ex.* II, *D.*—The last division of Ex. II is made up of examples of varying interest and importance. They are given here chiefly in order to prevent the pupil from acquiring an unduly narrow conception of the scope of graphic methods. The chief value of **Nos. 26** and **30** is that they introduce the pupil to curves, very different from the circle, yet having just as definite an individuality. **Nos. 27** and **28** show how the

course of a curve may be defined, either partially or entirely, by angular "co-ordinates". The solution of **No. 27** is given in fig. 3, that of **No. 28** in fig. 4. **No. 29** (solved in fig. 5) is an example of a type which always proves very interesting to beginners and has an obvious value as foreshadowing some

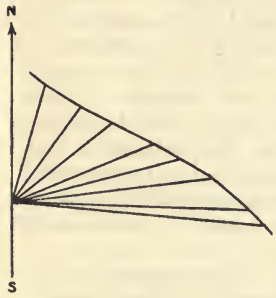


FIG. 3.

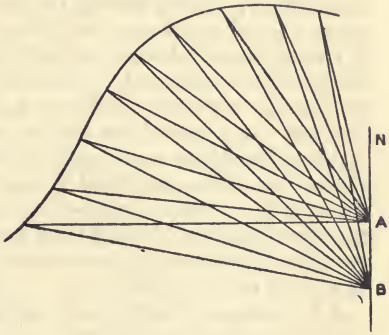


FIG. 4.

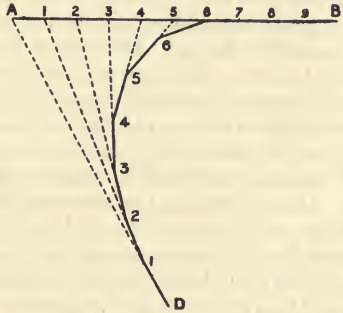


FIG. 5.

ideas of importance in the calculus. It is suggested in *A Rhythmic Approach to Mathematics*, the little book in which Miss Somervell has illustrated some of Mrs. Mary Boole's interesting ideas. Let AB be the path of the cyclist and the points on it marked 1, 2, 3, . . . the positions which he occupies at the end of the first, second, third, . . . intervals of time. Let D be the point from which the dog starts.

Then we may suppose that during the first interval the dog runs in the direction DA and covers the distance D1 which is, by hypothesis, half as long again as A1. At the end of the first interval we may suppose the dog to observe his master's change of position and to amend his own direction accordingly. Thus during the second interval he will run along 11 and will cover the distance 12. Again he changes his direction and during the third interval runs along the line 22, reaching the point 3 at the end of it. The process is continually repeated until the dog is found to have caught the man up. Lastly we argue that our solution of this artificial problem needs only slight amendment to fit the given case. The polygonal figure D123 . . . indicates a continuous curve which may easily be accepted as representing the actual path of the dog if he adjusted his direction at every moment to the varying position of his master. Such a curve is a "Curve of Pursuit".

§ 6. *Certain Principles of Method.*—The examples of Ex. II all conform to at least one principle which should never be contravened. It is that a graph should not be drawn unless there is a clear purpose to be served by it—a purpose that the pupil can understand and accept. "A train is going at 30 miles an hour. Determine by a graph how far it will go in 3 hours." This is a type of "graphical work" which offends the common-sense of the pupil and prejudices him against its legitimate uses. He gets to regard it as merely a capricious, inconvenient, and uncertain way of treating problems which can be solved with ease and certainty by ordinary arithmetic. He entertains the same sound objection to solving by graphs equations which yield without difficulty to algebraic methods. If there is nothing to be gained by it he sees no reason why the thing should be done.

The ends to be achieved by drawing a graph will, of course, vary greatly from case to case. From the logical point of view the simplest cases are those in which it serves as a "ready reckoner"—either by presenting a number of numerical facts in a form convenient for inspection or by solving troublesome calculations by a mechanical device. In most other cases the special service it renders is to disengage from a set of numerical data knowledge which is not obtainable from the numbers considered separately. The extent and value of this knowledge also vary considerably. The simplest

case is typified by the records of the snails' movements in Ex. II, No. 1, or the column-graph of No. 21 before the addition of the continuous curve. Such records are useful because the facts can be "taken in" from them as a whole more readily than from a column of figures. They give a clearer impression of how the snails' speed varied or the area of the flood increased—and one more easily retained in memory. The graph performs a much higher function when, as in No. 4, No. 22, and in most of the other examples of the exercise, it is used to bring to light the mathematical law underlying a set of data. It is important that clear thinking —graduated in "rigour" to the age and experience of the pupil —should accompany this use of the graph, otherwise there is some danger that a fine heuristic instrument may be perverted to the encouragement of slip-shod intellectual habits and the blunting of the logical sensibility. We have seen that graphs can in these cases be regarded as generalizations (§ 1). It must be noted, however, that they are not generalizations based upon analysis, but belong to the inferior type which requires a number of instances (ch. I., § 2); it is impossible to construct a graph upon one result. The trustworthiness of a graph depends, therefore, in the first instance upon the number and variety of the data originally plotted. Thus in Ex. II, No. 7, if the sun had been hidden by clouds from 10.20 to 12.30 the drawing of the middle part of the graph would have been attended with much uncertainty. It would have been unsafe to adopt any very definite conclusion about either the time of noon or the minimum length of the shadow.

But while the risk attending these graphic generalizations should always be clearly presented it should not be exaggerated. After all it is of precisely the same character and degree as the risk run by a physicist or a chemist who publishes a new law after carefully examining a number of well-chosen instances. He predicts that all other instances will be found to follow the law which he detects in his data. The confidence of the mathematician in a generalization not based upon analysis rests on the same foundation as that of the man of science—a belief in the prevalence of simplicity and continuity among natural phenomena of all kinds. "*Simplex sigillum veri*"; "*natura nihil per saltum facit*". Return in this connexion to Ex. II, No. 4, and let a, b, c, e, f be the terminal points of the lines which represent the recorded lengths of

the tulip. Then the problem before the pupil is not merely to join these points by a smooth curve—a problem soluble, of course, in an infinite number of ways—but to find the continuous curve which in addition to passing through *a*, *b*, *c*, etc., satisfies certain other definite conditions. Some of these conditions represent actual knowledge. For example, it may be taken as certain that the height of the tulip will not rise and fall rhythmically, so the graph must rise continuously from left to right. Other conditions are the expressions of more or less reasonable assumptions. Thus, although a plant may show rhythmic variations in its rate of growth, yet in the case of observations of a plant which cover a short period at the beginning of its career it may safely be assumed that the acceleration which is a characteristic of the successive weeks' growth when recorded was also a feature of the growth during the vacant fortnight. This assumption, translated into graphical terms, implies that the curve must show no "waves". Thus in the end ambiguity in the solution of the problem is reduced within very moderate limits. It has already been shown how similar assumptions limit the solution of problems like that of No. 22. From the point of view of the logical training of the pupil the important thing is that he should recognize that he is making certain assumptions and should understand how they work out in graphical terms. Fortunately the investigation of such matters adds not difficulty but interest to the lesson.

The same principle applies in a modified form in instances of which Ex. IV, No. 9, may be taken as typical. Here the pupil is to draw a graph in order to find by interpolation the square roots of numbers which cannot be determined by inspection. He plots the square roots of 1, 4, 16, etc., against the numbers themselves and draws a smooth curve through the points thus defined. This curve he uses to find the square roots of intermediate numbers. It should be noted that in this case there is no knowledge or plausible assumption from which we can deduce that the curve must be "smooth". The choice of the smooth curve is determined merely by its simplicity and uniqueness. Of all possible continuous curves through the given points this is the one which it is reasonable to try first. But in this case—as distinguished from those hitherto considered—we already know the law of which the graph is to be the expression. We can therefore use it to

test the success of our venture. Each pupil chooses at random one or two numbers within the range of the graph, and reads off the numbers which the graph asserts to be their square roots. The truth of this pretension is verified by multiplication. It is now impossible to doubt that though our initial choice of a smooth curve might have proved erroneous yet as a matter of fact it has been justified at least to the extent that square roots deduced from it may be expected as a rule to be accurate within the degree obtainable by inspection of the graph.

§ 7. *Graphs in Practical Work.*—When the points which are to determine the course of a graph represent the results of practical measurement the problem of drawing the appropriate curve is complicated by other considerations. Measurement is always subject to error, and the uncertainty about the correct position of the graphic points may, for this reason, be great enough to make it doubtful how the curve should be drawn even when its general form is already known. When the general form of the graph is itself unknown the draughtsman will attach much weight to considerations of simplicity. Thus if the assumption that the graph is "meant to be" a straight line is not grossly discordant with the actual position of the points, he will adopt it in preference to the assumption of a more complicated curve. But when he has learnt or has assumed that the graph has a certain general form he has still to determine the particular instance. In advanced work it is customary to apply in this connexion the "method of least squares" which is illustrated in the last section of this volume. In elementary work no systematic method of dealing with the difficulty is generally used. The pupil is told to select the straight line or curve which his eye judges to fit the points best. In Ex. XXVI, D, a simple method is discussed for which there is a good deal to be said on statistical grounds, while it offers at least a definite principle for dealing with the kind of situation which is now in view. It may be illustrated by No. 17 (*Exercises*, p. 144). The solution is shown in fig. 6 in this book. Since it is known that P is proportional to Q^2 the eleven given values of P are plotted (in the left-hand portion of the figure) against numbers which represent the squares of the values of Q. Each of the graphic points corresponds, of course, to a definite possible value of the ratio P/Q^2. This value would be the tangent of

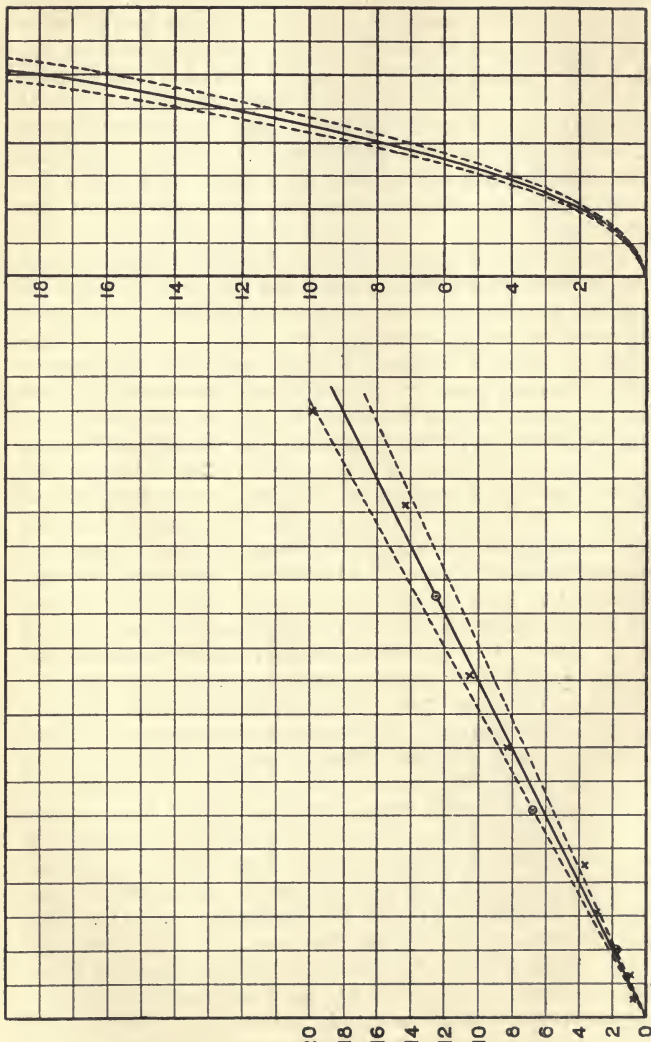

0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36
0 2 4 6 8 10 12 14 16 18
0 2 4 6 8
0 2 4 6 8
2 4 6 8 10 12 14 16 18 20

FIG. 6.

the angle between the horizontal axis and the line joining the graphic point to the origin. If all the points were joined up with the origin the resultant lines would constitute a fan of 11 rays. In the figure the middle of these rays (the "median") is shown in a firm line, while the third from each end (the "quartiles") are shown in dotted lines. The others are omitted for clearness. Now let us suppose that the correct value of the ratio is k and let us, assuming some numerical value for k, find the deviations from it of the ratios represented by the points in the graph, and add them together. Then it can be proved with little difficulty [1] that the sum of the deviations can never be less than when k is assumed equal to the ratio represented by the median line. In this very definite sense, it can, therefore, be said that the median represents best the general effect of the observations. But it is clear that, on another occasion, while the median line might fall in the same position as in the present instance the fan as a whole might exhibit important differences. It might be more compact or more extended. The former case would imply greater, the latter less concordance among the measurements than in the present instance. Thus it is evident that the median should be supplemented by some indication of the degree of "scatter" among the individual observations. But this is a function which the quartile lines are well suited to perform. For it is evident that if, speaking generally, the graphic points deviate widely from the median there will be a wide angle between the quartiles; while at the same time their position will not be affected by the accident of a single "wild" observation. Thus the lines presented in the left-hand part of fig. 6 may be taken as summarizing in a simple and useful way the information conveyed by the original eleven observations. On the right of the figure this summary is translated into the graphic form appropriate to the description of a case in which one variable is proportionate to the square of another. The abscissæ are now the values, not of Q^2 but of Q; the curve drawn with a firm line is the semi-parabola corresponding to the median value of the ratio, while the dotted curves correspond in the same way to the quartiles.

§ 8. *The Graph in Algebraic Theory.*—Our discussion has as yet hardly touched upon what is after all the main function

[1] See *Exercises*, II, *Ex.* CXXI, No. 17.

of the graph in algebra. This, as described in § 1, is either to point the way to a formula by which it may itself be superseded or to unfold more fully the implications of a formula whose properties have been only partially explored. On the other hand this function is illustrated so constantly throughout the book that it will suffice to give here a very brief review of the different uses which are included in it.

The first of these is well illustrated by the lessons and examples which deal with the familiar identities $a^2 - b^2 = (a + b)(a - b)$, $(a + b)^2 = a^2 + 2ab + b^2$, etc. (chs. VII., VIII., IX., Exs. V-XI). In these the figures composed of squares and rectangles are used, much as the Greeks used them, to reach certain analytical results of importance. The only important difference between the older usage and the present one is that in these lessons the manipulation of the graphic forms serves a purely temporary purpose. It is used to suggest algebraic identities which are afterwards established upon an independent basis. The earlier lessons in Section III (Logarithms) illustrate a similar use of the graph. Here the "growth-curves" of *Exercises*, fig. 50 (p. 273), are employed temporally for the solution of problems which are afterwards to be solved by logarithms, and the properties of the curve are made to suggest the algebraic method which eventually makes a graphic method unnecessary.

The second use of the graph in algebraic theory is exemplified every time the student draws the graph of a "function". The mere statement that two variables are connected by the relation $xy = a$ conveys comparatively little of its implicit contents to a person who has not explored them by means of graphic analysis. Even students of considerable mathematical knowledge may often experience the truth of this observation; for example, when they try to realize the figure which corresponds to a simple function involving three variables. In this kind of application the graph not only serves to bring out particular features of the function which might otherwise not be noticed or not be realized; it also serves as a kind of challenge to the student to complete or supplement his view of the function. Thus at the lowest the drawing of a graph gives a useful occasion for a systematic and orderly examination of the properties of the function. But it may often do much more. Thus the exercises of Section VII (Complex Numbers) illustrate important cases in

which the attempt to give a graphical representation of a function leads to the discovery of a whole field of values which might otherwise have remained unknown.

Lastly it must be observed that a graph is more than a means by which a given function can be conveniently explored. It is also by far the best means by which, in the earliest stages, the function is taught and symbolized. In fact, long before the student is ripe enough to discuss linear, parabolic or other "functions" under those abstract names his mind has been prepared for the reception of the notions they connote by the repeated emergence of straight lines and parabolas in his graphic exercises. And as the graph is the forerunner of the idea of an algebraic function so it remains the chief support of that idea and the chief means of maintaining in the student's mind the distinct individuality of the different functional relations. For this reason each of the more important functions is taught in these lessons in close connexion with the curve which serves as its graphic symbol, preserving the outlines of the function distinct in the memory and ensuring the certainty of its recall.

§ 9. *The Graph and Co-ordinate Geometry.*—The discussion of this chapter should have made clear the essential distinction between the use of the graph in algebra and the formal study of co-ordinate geometry. Briefly summarized, the difference is that in algebra graphs are studied only for the light they throw upon the properties of functions, while in co-ordinate geometry the algebraic properties of functions are of interest only in so far as they contribute to the exploration of the properties of curves. This distinction between the two points of view is important and should be used to define the range of the geometrical ideas and knowledge to be embodied in the algebra course. Although it would be pedantic to insist upon a too scrupulous observance of the frontier, yet there is much to be gained in definiteness of aim and unity of method by restricting the attention of the student in this course to the graphical work which has direct relevance to algebraic investigation. It is well to add that these observations apply to graphs of functions involving three variables as directly as to those involving two variables. The surfaces which correspond to the former should be treated—as they are in Section IV, Ex. LXXVII—simply as three-dimensional graphs, and should be studied in the algebra course for the same reasons

and to the same extent as graphs which can be represented upon squared paper.

§ 10. *Some Practical Suggestions.*—This chapter may be brought to an end by a few practical suggestions which may be of use, at least to younger teachers. The pupil should be taught to recognize that a graphic record is useless unless its object is clearly stated and unless the scales of representation are properly labelled and graduated. Beginners are prone to label the vertical scale in such a way that the feet of the letters are (like those of the horizontal scale) towards the edge instead of towards the middle of the paper. A graph gives excellent opportunities for insisting upon the "great school virtues" of neatness, carefulness and accuracy. The finest and cleanest of lines drawn with the sharpest of pencils should be demanded, and no obscurity or untidiness of figures or lettering allowed. The pupil should be taught to choose the scales of representation so as, on the one hand, to make the work of plotting and reading points easy, and on the other hand, to utilize the sheet of paper as completely as possible. The natural boy and girl tend to produce a graph which clings closely to one edge of the paper. They should be shown that such graphs are wasteful both of paper and of the possibilities of accuracy.

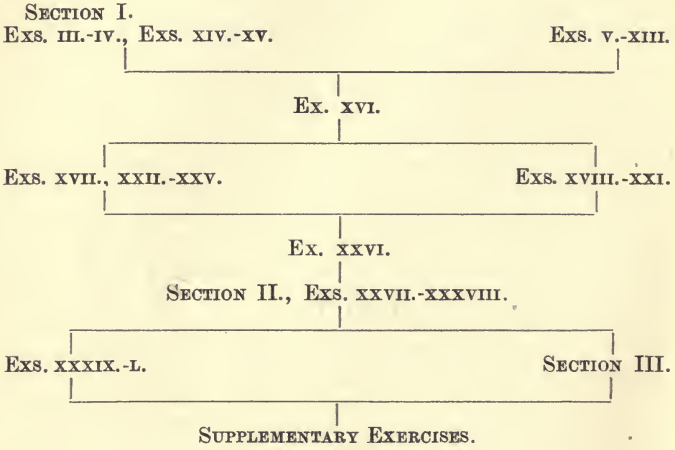
The great extension of graphic work in schools has led to the production of quantities of inferior squared paper. The cheaper kinds are badly inaccurate. They should be carefully avoided, for their use militates directly against the teacher's attempt to insinuate ideals of painstaking and accurate work.

PART I.

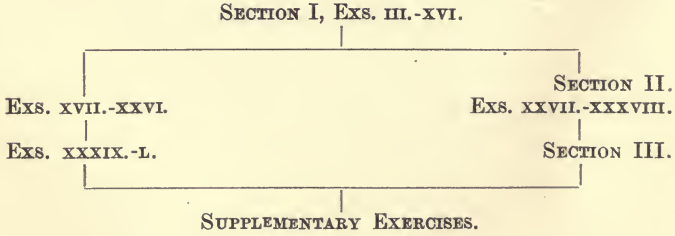
SCHEME OF ALTERNATIVE ORDERS OF STUDY.

(Cf. p. 59.)

A.



B.



CHAPTER V.

INTRODUCTION TO PART I.

§ 1. *The Essentials of a Course in Algebra.*—The exercises of Part I are intended to cover, and to be confined to, those portions of the subject which are of such fundamental importance that their study should form part of every scheme of secondary instruction. When the details of such a curriculum are worked out there is necessarily found to be scope for much difference of opinion among those qualified to speak. Any particular set of proposals will almost certainly exhibit —by inclusion, rejection and emphasis—features for which general acceptance can hardly be expected. Nevertheless there is among thoughtful and experienced teachers a growing agreement about the essentials of a general course in algebra, and unanimity of opinion that the prevailing tradition fails in some important respects to embody them. It is, no doubt, an advantage that within the broad lines of this concensus there should be many different individual presentations of the subject.

The presentation offered in this book postulates that the course to be taken by all boys (and possibly by all girls) in a secondary school shall consist of the following items :—

(*a*) The use of the formula as a means of making and of expressing arithmetical generalizations, and of describing the quantitative regularities which characterize physical, social and other phenomena. The making of formulæ, including practice in the simplest forms of algebraic symbolism. The interpretation of formulæ and the determination of particular results by substitution.

(*b*) The art of graphic representation. The uses of the graph as a subsidiary instrument of analysis and generalization.

(*c*) The manipulation of formulæ in order to bring out the further relations which a given generalization may imply. The application of these processes to the solution of problems of real interest and of practical importance.

(*d*) In particular, the use in formulæ of the symbols of the trigonometrical ratios, the manipulation of such formulæ, and their application to simple practical problems.

(*e*) The extension of arithmetical ideas to include the complete scale of positive and negative numbers. The use of these positive and negative numbers in formulæ (including trigonometrical formulæ) and the rules to be observed in manipulating them.

(*f*) The use and theory of logarithms.

(*g*) A simple introduction to the ideas and methods of the integral and differential calculus.

(*h*) The idea of a "function" as a generalization from concrete instances of numerical dependence. The simpler functions of a single variable and their characteristic graphic symbols.

The exercises in which this programme is developed are grouped into three sections. Speaking roughly, Section I ("Non-directed Numbers") covers (*a*), (*b*), (*c*), and (*d*); the items in (*e*), (*g*) and (*h*) fall mainly into Section II ("Directed Numbers"), while those in (*f*) are dealt with in Section III. In addition a group of "Supplementary Exercises" reviews the whole course and completes the development assigned to the various topics.

§ 2. Section I. *Non-directed Numbers.*—The course outlined in Section I, Exs. I-XXVI, is intended to give the pupil clear ideas about the fundamental processes of algebra together with sufficient technical facility to ensure his appreciation of their value as instruments of mathematical statement and inquiry. The fundamental processes are taken to be: (i) the making and use of formulæ; (ii) factorization; (iii) the converse process of expanding a product; (iv) the simplification of easy algebraic fractions; (v) the process here called "changing the subject of a formula" and commonly known as "the solution of simple equations". The programme includes in addition: (vi) the fundamental ideas of functionality, illustrated by a simple study of direct and inverse proportion. The use of the trigonometrical ratios is taught in connexion with this study.

The title "Non-directed Numbers," is meant to imply that in this section algebraic symbols always stand for the ordinary, signless numbers of arithmetic. Thus if, in a given instance, the symbol a represents the number 8, it means neither $+8$ nor -8, but simply the cardinal number 8 used in the sense intended when I say that I have eight coins in my pocket or that there are eight books upon the table. We might have said, reviving an ancient term, that the subject of Section I is "specious arithmetic". Not until Section II is reached are symbols to be regarded as standing for numbers accompanied by signs, such as $+8$ or -8. There are two good reasons for this procedure. The first is that the expression of generalized statements by formulæ and the use and manipulation of numbers accompanied by signs are two distinct processes which have no necessary connexion with one another. Each has its own difficulties for the beginner, difficulties which are best overcome if faced separately. The second reason is that the two processes are not only different in kind; they depend upon the presence of characters in the child's mind which begin to ripen at different ages. It is easy enough to make a boy or a girl apply mechanically the rule that "like signs produce *plus*, unlike signs *minus*"; but if the rule is to be used with intelligence the teacher must be able to appeal to logical powers and interests which have rarely emerged at the age when lessons in algebra begin. Observation will, in fact, show that the pupil who has been early taught the properties of positive and negative numbers rarely uses them spontaneously in his thinking. His mind works freely only among signless numbers with their familiar properties. This fact is itself a strong indication that numbers accompanied by signs have been taught prematurely.

It is of great importance to understand the bearing of these observations upon the "rule of signs". There are, strictly speaking, four distinct types of problems covered by this rule. The first is the type discussed in Ex. X, note to No. 5, and in ch. IX. B, § 3. It is there shown (for example) that

$$\begin{aligned}\text{any number} - 5 \times 6 &= \text{the same number} - (9 - 4) \times 6 \\ &= \text{the same number} - 9 \times 6 + 4 \times 6.\end{aligned}$$

The change of sign that occurs here is easily justified by purely arithmetical considerations: taking away 6 fives is equivalent to taking away 6 nines and adding 6 fours. The next two types are those represented by such symbolisms as :—

$$(-3) - (-7) = -3+7$$
and $$(-3) \times (-7) = +21$$

discussed respectively in ch. XVIII. B and C. The fourth type is exemplified by the process of "multiplying" two factors, such as $a + b$ and $a - b$, when the literal symbols represent positive or negative numbers. In none of the last three cases does *plus* mean "add" or *minus* "subtract" in the primary senses of those terms. The appropriate "rule of signs" must therefore be discussed afresh in each of them. The fact that as the result of this discussion the same rule is found to hold good in all four types does not make it a less serious crime against logic to *assume* that the rule which obtains in the first case must be valid in the others. Yet this practice, it is to be feared, is far from unusual.

In Section I, then, symbols always imply numbers without sign; *plus* and *minus* have their direct arithmetical significance; a number is never "subtracted" except from a number larger than itself; and the rule of signs is considered only in relation to the first of the four cases just distinguished.

The exercises fall naturally into two groups—Exs. I-XVI, and Exs. XVII-XXVI. Those of the first group teach the fundamental uses of formulæ and the main types of algebraic manipulation—factorization, the simplification of fractions and the processes by which the "subject" of a formula may be changed. In the second group the formula describing the numerical relations between a particular pair of variables ceases to be the centre of interest, the pupil's attention being now directed to the resemblances which are so often exhibited in the numerical relations of pairs of variables of widely differing character. Only the simplest and most important of these resemblances are studied, namely, those to which the mathematician applies the terms "direct" and "inverse proportion". It has already been indicated that the beginnings of "numerical trigonometry" have their place here, the properties of the tangent, sine and cosine being treated as especially important instances of direct proportion.

§ 3. SECTION II. *Directed Numbers.*—In the strictest sense all algebra is "generalized arithmetic"; that is to say, all algebraic processes are based ultimately upon the properties of numbers, and the results of algebraic investigation have their meaning and value solely in the fact that they are discoveries or demonstrations of further general or specific properties

of numbers. Nevertheless there are important differences in nature between the ground covered in Section I and that of Section II. In the first place the numbers considered in Section I are, as we have seen, the signless numbers of elementary arithmetic—the numbers which begin with zero and are continued through 1, 2, 3, etc., in an endless series. The generalizations of Section II are based upon a different series—the series of "directed numbers," or numbers *with* signs, which has neither beginning nor end but from any starting-point (such as zero) can be continued forwards and backwards without end. In the second place this new series differs from the numbers of ordinary arithmetic in its origin. The signless numbers were known and used long before algebra was invented; the use of numbers with signs is itself a product of the development of algebra. It is, in fact, the most important instance of the characteristic explained in ch. I., § 4. Consider a simple example: I am travelling with velocity v towards a point from which at the present moment I am at distance d_o; what will be my distance from it (d) after time t? The formula $d = d_o - vt$ sums up the solutions of all particular problems of this form which it is possible to state. But, regarded as a "shorthand" prescription of certain operations to be carried out with given numbers it has a strictly limited field. The instructions which it gives can be fulfilled only if vt is not greater than d_o. Now the use of numbers with signs enables us to set this limitation aside and to obtain results from the prescribed operations even when vt is greater than d_o. Moreover it is easily shown that such results are solutions of actual problems akin to, if not identical with, the original ones. (In the present case a negative result is the answer to the question: "How far shall I be *beyond* the point in time t?") Thus the introduction of the new number series kills two birds with one stone. It removes our natural dissatisfaction with a generalization whose scope is limited differently by the conditions of every different problem; and it greatly increases the range of problems to which the single formula applies.

The study of the consequences and advantages of adopting numbers with signs as the basis of algebraic generalizations constitutes, then, the main subject of Section II and gives unity to the topics considered. First the new numbers are introduced and the laws of their combinations are determined

(Exs. XXVII-XXIX). Subsequent exercises illustrate the extraordinary increase in scope and power which accrues to the notations, notions and processes of Section I—formulæ, identities, "change of the subject," trigonometrical ratios, specific functional relations—when the symbols are taken to describe numbers with signs instead of the numbers without signs of ordinary arithmetic.

Upon this plan an expression such as $a + b$ may refer to very diverse numerical combinations: for example, $+8+3$ (if $a=+8$ and $b=+3$), $+8-3$ (if $b=-3$), $-8+3$, etc. Similarly the expression $\times 10^n$ may mean a number either of successive multiplications or of successive divisions by 10 according as n is positive or negative. It follows that throughout Section II much more attention must be paid to algebraic *form* than in Section I. This statement does not mean that the practical and heuristic aims of the former section are abandoned. It means merely that a scientific study of algebraic form is now recognized as necessary to their successful pursuit. In other words, the idea of a *function* receives considerable emphasis, especially in the latter half of the work.

Like those of Section I the exercises of this section fall naturally into two groups—Exs. XXVII-XXXVIII and Exs. XXXIX-L. There is indeed a rough correspondence between the divisions in the two sections; for the aim of each group of exercises in Section II may be regarded as that of working out the results which follow when "directed" are substituted for "non-directed" numbers in the arguments of the corresponding group in Section I. This general statement applies to one particularly important instance—namely to the development of the ideas and methods which are generally regarded as belonging to the differential and integral calculus. The starting-point of this development will be found in some of the earlier exercises of Section I. Thus in Ex. IX the notion is introduced of an "approximation-formula," i.e. of a formula which gives results true to a certain degree of accuracy when one of the magnitudes is small relatively to others. In the last exercises of Section II the notion of an approximation-formula is generalized into the idea of a "differential formula" which may be applied in accordance with definite rules to functions of a definite form. Again in the first division of Section II (Ex. XXX) the young student learns to solve a certain group of problems (arising out of the

study of arithmetical series) by a method which is, in effect, integration. At a later point (Ex. XLVII) this simple method is itself generalized, and the student learns that a large number of important problems can be solved by a knowledge of the rule which connects a function of given form with its "integral".

§ 4. Section III. *Logarithms.*—The best way of approaching logarithms is a subject upon which teachers of mathematics hold very different opinions. The quarrel revolves about the question whether logarithms should be taught before fractional indices or fractional indices before logarithms. The former method follows the course of history, the latter complies with the tradition of the text-books. The present writer for some time dallied with a method which sought to combine the advantages of both modes of attack. After further experiment and reflexion he has become convinced that a treatment which keeps close to the historical evolution of the subject is on the whole the simplest and the most satisfactory from the logical point of view. The arguments adduced against a too early introduction of directed numbers weigh also against a treatment of logarithms based upon the theory of indices. It is at least doubtful whether the somewhat subtle logic needed in this case is really appreciated by the young pupil. On the other hand a treatment which follows the reasoning of Napier, Briggs, Mercator and Gunter is not only much simpler but also means much honester thinking on the part of the student. In Section III, then, the reader will find the usual relation between logarithms and fractional indices inverted. The theory of the latter is based upon the theory of the former.

Though this order of presentation follows the order of emergence in history of the ideas of logarithms and fractional indices it does not represent exactly the manner in which the latter notation found its way into algebraic practice and theory. Fractional indices were actually invented by John Wallis in the course of his investigations into the "arithmetic of infinites".[1] The method of treatment illustrated in Ex. LVIII is a kind of idealization of the course of history; it represents what might have happened rather than what actually took place. It may perhaps claim to be more in

[1] See the present author's articles on Wallis in the *Mathematical Gazette*, Dec. 1910 and Jan. 1911.

accordance with the modern tendency to "arithmetize mathematics" than the traditional method of the text-books. Nevertheless, as the latter illustrates important points in the general theory of algebra it is offered to the maturer student in the first section of Part II.

The exercises of Section III depart from history in another important respect. Napier invented logarithms in order to lighten the labour involved in the long multiplications and divisions of spherical trigonometry. The disadvantage of adopting this starting-point in teaching is that for some time it is easier for the student to perform multiplications and divisions by arithmetic than to find the results by logarithms. If the numbers are large enough to make the arithmetical operations troublesome they are too large to be dealt with satisfactorily by the four-figure logarithms with which the student usually begins. It seems better, therefore, to commend logarithms to the beginner by showing their application to problems which could not be solved by ordinary computation. In accordance with this view the earlier exercises of Section III are given to the study of "growth problems" in which a magnitude is contemplated as increasing or decreasing in accordance with the geometric or "compound interest" law. Problems of this kind, involving fractional periods of time, cannot well be solved except by logarithms or by graphic methods which lead directly to the conception of logarithms. Such a method of procedure has the further advantage of directing attention naturally to important financial and social phenomena in which the logarithmic and "exponential" functions are exemplified. Thus it becomes possible, in the last exercise of the section, to introduce the student to that important entity "e" in circumstances which are calculated to give him sound ideas about its real significance.

§ 5. *Supplementary Exercises.*—The supplementary exercises may be regarded as an appendix to Sections II and III, developing still further the topics of those sections and, in some cases, bringing them into relation with one another. Their position in the book is meant also to imply that their contents have not the same fundamental importance as those of the earlier sections. They may without great harm be reserved, wholly or in part, for brighter or older pupils. In Ex. LXII the "sum and difference" formulæ of the trigono-

metrical ratios are investigated. In Exs. LXIV and LXV these are used to complete the earlier study of functions of the second degree and their graphic forms. Exs. LXVI and LXVII supplement certain statistical notions acquired as far back as Ex. XXVI, and (taken together with Ex. LXIX) complete the elementary treatment of the calculus by exemplifying some important applications of its methods. Ex. LXVIII collects a number of results scattered through previous exercises and generalizes them into the binomial theorem.

§ 6. *Order of Study.*—The teacher may welcome some suggestions about the order in which the exercises should be taken. The schemes printed on p. 50 suggest two alternative orders each of which has certain advantages. It is assumed in each that two weekly lessons are given to algebra. The first scheme is probably the more suitable for a class which begins the book at a fairly early age and with no previous knowledge of the subject. It subdivides the exercises in the first group of Section I upon a plan which carries the pupil quickly from the first lessons upon the formula to the lessons on "changing the subject"—i.e. to simple equations. Factorization, approximation-formulæ and the simplification of easy fractions are, upon this plan, studied in a parallel series of lessons. A similar subdivision is made in the second group of Section I. This has the effect of bringing together, on the one hand, the exercises which deal with the various types of proportionality, and, on the other hand, the exercises which cover the first year's programme in "numerical trigonometry".

Scheme B is more suitable for those who begin the study of algebra at a later age or with previous knowledge of the earlier parts of the subject, and may also be followed by teachers who wish to reach as quickly as possible the doctrine of positive and negative numbers. It differs from the former scheme in prescribing the first group of exercises in Section II to be taken side by side with the second group of Section I. The first group of Section I may be taken either *seriatim* or upon the plan indicated in Scheme A. It will be observed that according to both arrangements Section III (Logarithms) is to be worked simultaneously with the second group of Section II. Reference to the table of exercises on p. 158 will show that with this sequence the study of logarithms and fractional indices follows shortly after the consideration of positive and negative indices in Exs. XXXIII and XXXIV.

§ 7. *Public Examinations.*—Speaking generally the course of Part I covers, and indeed exceeds, the requirements of school-leaving examinations and tests of similar standard. The teacher should, however, note that, for reasons assigned, certain topics, often included in the syllabuses of these examinations, are omitted. There is no treatment of the "imaginary" roots of equations nor of permutations and combinations in algebra, nor of circular measure and the ratios of angles of unlimited magnitude in trigonometry. The author hopes and believes that in these exclusions he has merely anticipated the action of the examining authorities. Meanwhile the teacher who deems it necessary to deal with these topics in the general course may easily supplement Part I by the few exercises in Part II in which they are treated.[1]

[1] This is, perhaps, a suitable place for a note of acknowledgment. The author has long recognized the importance of distinguishing clearly between the algebras of "non-directed" and "directed" numbers. In publishing a scheme of work based largely upon this distinction he has, however, been anticipated by Messrs. Barnard and Child in their *New Algebra*. The pedagogical ideas embodied in the *New Algebra* appear in some respects to be very different from those of the present author; he is the more glad, therefore, to confess his admiration for a work which exhibits so much ability and sincerity.

SECTION I.

NON-DIRECTED NUMBERS.

THE EXERCISES OF SECTION I.

*** The numbers in ordinary type refer to the pages of *Exercises in Algebra*, Part I; the numbers in heavy type to the pages of this book.

EXERCISE		PAGES
I.	The "Shorthand" of Algebra	1, 25
II.	Graphic Representation	4, 31
III.	The Writing of Formulæ	13, 63
IV.	The Reading and Use of Formulæ	23, 67
V.	Factorization (I)	32, 82
VI.	Factorization (II)	39, 87
VII.	Square Root	44, 90
VIII.	"Surds"	46, 93
IX.	Approximation-Formulæ (I)	48, 72
X.	Approximation-Formulæ (II)	53, 72
XI.	Approximation-Formulæ (III)	55, 72
XII.	Fractions (I)	51, 96
XIII.	Fractions (II)	63, 98
XIV.	Changing the Subject of a Formula (I)	70, 104
XV.	Changing the Subject of a Formula (II)	74, 106
XVI.	Supplementary Examples	81, 79
XVII.	Direct Proportion	103, 117
XVIII.	The Use of the Tangent-Table	108, 121
XIX.	The Use of the Sine- and Cosine-Tables	112, 124
XX.	Some Navigation Problems	118, 129
XXI.	Relation of Sine, Cosine and Tangent	120, 132
XXII.	Linear Relations	122, 136
XXIII.	Inverse Proportion	126, 145
XXIV.	Proportion to Squares and Cubes	129, 149
XXV.	Joint Variation	132, 154
XXVI.	Supplementary Examples	135, 114

CHAPTER VI.

THE PROGRAMME OF SECTION I (EXS. I-XIV).

§ 1. *The Cultivation of the Formula.*—Formal work in algebra—as distinguished from the incidental use of symbolism in arithmetic and elementary science (ch. III.)—is here planned to begin with lessons intended to cultivate the formula as an instrument of mathematical statement and investigation. When it is considered how essential is their use in a vast range of trades and professions—from plumbing to *Dreadnought* building—it is hardly extravagant to say that facility in the working, interpretation and application of formulæ is one of the most important objects at which early mathematical studies can aim. A beginning at this point secures, therefore, the tactical advantage of giving the pupil his first view of the subject on its most obviously useful side.

The cultivation of the formula involves four distinct elements : (*a*) practice in analysing arithmetical processes and rules of procedure ; (*b*) practice in symbolizing the results of analysis ; (*c*) practice in interpreting given pieces of symbolism ; (*d*) practice in " substitution ". The first two constitute the art of formulation ; the second two the art of using formulæ. It is advantageous to give separate study to these two sides of the work. For this reason Exs. I and III are devoted almost entirely to formulation and Ex. IV to the reading of formulæ and to substitution. It is not intended, however, that the whole of Ex. III should be worked before Ex. IV is begun. After a good start in formulation has been made the two exercises should be carried on concurrently.

§ 2. *Formulation* (Exs. I, III).—Ex. I contains very simple examples of formulation. We have seen that it is to be regarded as representing the ground covered by the preliminary work described in ch. III. Where the plan there recommended has been followed, Ex. I will serve as revision ex-

amples preparatory to the formal study of the subject. In this case the exercise may well be taken orally or set for homework, with or without preparation in school. If the simple uses of formulæ have not yet been taught Ex. I should be worked carefully according to the methods of ch. III.

These methods hold good equally for Ex. III. That is, the object in each of the examples is (i) to formulate *in words* the mathematical relation or the rule of procedure which the example illustrates,[1] and (ii) to reduce this verbal expression to a symbolic or "shorthand" expression. The quantity represented in symbolism on the left-hand side of the sign " = " may conveniently be called the "subject" of the formula. Thus, in the first of the formulæ in the footnote "the tangent of the angle a" is the subject; in the second formula, "the volume of a cylinder". The term is useful for at least two reasons. It reminds the pupil that his formula is always a symbolic rendering of a verbal statement; and it helps the teacher to exact precision in verbal formulation before the symbolic transcription is made. The teacher should never neglect this opportunity of cultivating accuracy and directness of statement.

In some examples (e.g. Nos. 21, 22) before a formula is demanded a numerical instance is given. This is, of course, intended to help the pupil to formulate the essential rule. In other cases where he cannot formulate the rule without such an instance he should invent one for himself. The numerical example should not be worked out to its result but should be set down in such a way that all the steps involved in obtaining the result are clearly exhibited. As a rule these steps will be taken one by one without any clear consciousness of the plan of procedure as a whole. The pupil should then analyse the working so as to make himself clearly cognisant of the details of the general plan, apart from the particular numbers in which, in this case, the plan is realized. The

[1] A formula can usually be regarded as stating *both* a mathematical relation (i.e. a numerical identity underlying diverse equivalent forms) *and* a rule of procedure. In most cases, however, one of these ways of looking at it is more natural than the other. Thus the formula $\tan a = \sin a/\cos a$ suggests most readily a fact of relationship, the formula $V = \pi r^2 h$ a practical rule. Cf. the remarks on the sign of equality, ch. I. § 3.

formula is, as we have already seen, nothing more than a statement of this plan in accordance with a conventional system of symbolism.

The formula obtained by analysis is, of course, not a description of the special case as such (ch. I., § 2). The special case is used merely as a challenge to the student to formulate the procedure which he would be bound to adopt *in any case*. As he gains experience it should become less necessary to begin with a numerical instance. At each step in the evolution of the formula he should determine what he would do next *if he were concerned with a particular case*, and should write down, symbol by symbol, his statement of procedure without needing to have the special case as a whole before him. Finally he should cease to be conscious that he is making any appeal at all to numerical instances and should handle his symbols exactly as if they were figures [1] (ch. I., § 5). But it cannot be urged too emphatically that the best way to give the student this degree of mastery over symbolism is to allow him first to obtain the full value of the lower degrees.

The examples of Ex. III have been selected to give occasion for introducing, one by one, the simpler forms of algebraic symbolism. Thus they necessarily illustrate some of the most important types of numerical fact which the world presents to the mathematician for analysis. Our conception of algebra dictates that the pupil should have some real acquaintance with these facts in their concrete settings before he is asked to bring his analytical symbolism to bear upon them. It follows that the order of the examples is to some extent arbitrary; for it must depend partly upon the extent of the pupil's familiarity with their subject-matter. Thus in a school where practical mensuration and simple physical measurements have an early place many of the examples placed later will seem easier than their predecessors, because the class is familiar with the facts and processes to be analysed. The teacher should, therefore, consider the examples as a whole and should take them in the order most appropriate in the circumstances.

That the pupil shall always in the algebra lesson feel that

[1] Example: I buy an article for £a and sell it for £b. My gain is, therefore, $\frac{100\,(b-a)}{a}$ per cent.

he is face to face with something in the real external world and that his business is to give in symbolism an account of its behaviour in its numerical aspect—this principle is of quite fundamental importance. The teacher should keep it prominently before him at every stage. Thus whenever the facts underlying an example in Ex. III are unfamiliar something should be done to make them real to the class before the analysis is attempted. In some cases the exhibition of a model or a piece of apparatus will be possible and is the best means of producing understanding. In other cases the same result may be reached by putting the bare facts in a pictorial setting. For example, finding his boys baffled by the relatively abstract statement of **No. 3**, the teacher may speak to the following effect: You are on the side of a hill and you notice a number of policemen, one at the top and others at different points in the hedge along the road. Suddenly a motor car appears on the summit of the hill. The policeman stationed there looks at the car and at once writes something in his note-book. The car descends the hill, of course with increasing speed. Exactly a minute after its first appearance it passes the second policeman, after another minute the third, and so on. Each policeman makes an entry in his note-book. Afterwards they gather together and you hear what they have noted down. The first one says, "He passed me at 14 miles an hour"; the second, "He passed me at 17 miles an hour"; the third, "He was going 20 miles an hour when he passed me". What will you expect the fourth man to say if the rule shown by the first three statements continues to hold good? . . . the fifth? . . . the tenth? etc., etc.[1]

In conclusion the reader is reminded that an expression of the form $a - b$ implies throughout Section I that a smaller number is to be subtracted arithmetically from a larger. If the number intended by b becomes larger than the number intended by a the symbolism must be held to be no longer applicable. If the difference is still to be taken the operation of taking it must now be represented by the symbolism $b - a$. Ex. III, **Nos. 28** and **32** give instances of formulæ in which a reversal of the symbolism is required by a change in the

[1] This story is not offered as a veracious "word-picture" of a police-trap!

values of the numbers. In Section II (Exs. XXVIII, XXIX) it is shown that the device of using numbers with signs enables us to include these different cases in a single symbolic expression.

§ 3. *The Reading of Formulæ. Substitution* (Ex. IV).—The correlative of the power to express a verbal statement in algebraic symbolism is the power to retranslate such symbolism into verbal terms. Most of the examples of Ex. IV lend themselves to the cultivation of this power; some are specially intended to exercise it. It is a good plan to have a number of the formulæ "translated" round in class as the sentences of a Latin or French exercise are translated—the same care being taken to secure a ready and exact rendering.

But the chief object of Ex. IV is to cultivate the art of substitution. The principles of this art are simple and have been sufficiently discussed in ch. III., § 5. We pass on, therefore, to note that, as far as possible, the formulæ have been chosen for the interest of their subject-matter as well as for their value as exercises in substitution. It is not difficult to harmonize the two conditions, for the symbolic forms which have most frequent practical use naturally offer the most profitable field for exercise in substitution. Many of the examples in substitution have, therefore, been drawn from Molesworth's *Engineer's Pocket-Book* and similar formularies. In some cases (e.g. **Nos. 3, 6, 15**) the formulæ admit of simple practical applications. A few such applications will do more to illustrate the value of algebra than many formal lessons. The teacher should, therefore, make as much use of them as possible. Many more examples of this type could, of course, be given, but it is advisable to avoid anticipating formulæ which the pupil may encounter in his later studies in science, etc. This principle limits the selection to empirical formulæ and rational formulæ which lie outside the ordinary scope of school work.

Graphic methods enter in two distinct ways into the solutions of these examples. In some cases (e.g. **Nos. 13, 21**) the object of the graph is to give a conspectus of the particular numerical facts covered by the formula. Such a graphic presentment of the results of successive substitutions conveys a fuller meaning than can be gathered easily from the numbers themselves. Moreover, a graph based upon a sufficient number of results obtained by direct substitution can be used to

save the labour of further substitutions. This function of the graph has still more obvious usefulness when, except in certain cases, direct substitution is impossible. Thus in **Nos. 6, 7, 8,** it is easy to obtain results by substituting under the radical sign numbers, such as 16, 81, etc., which are perfect squares; but for boys who have not learnt the square root process other substitutions are far from easy. It is possible by trial to hit upon a number which when squared comes (for example) approximately to 75, but the process is inconvenient and uncertain. The class should, therefore, draw a careful graph in which the abscissæ are the numbers from 0 to 100 and the ordinates are their square roots. Points representing the square roots of 1, 4, 16, 25, etc., should first be plotted and a smooth curve drawn carefully through them. (Cf. ch. IV., § 6.) It is obvious that the square root of a number 100 times as large as a given number will be 10 times the root of the latter. For example, the square root of 2600 is 10 times the square root of 26. This principle will be needed again in Ex. VII.

The teacher should take occasion to point out that finding a square root is, like finding a quotient or guessing a riddle, an inverse process.

The arguments of ch. IV., § 6, apply to both the uses of the graph described in this section, for in each case we start out with a formula and are seeking its graphic expression. The curve based on a few calculations must be tested by confirming the results of random interpolations. If each member of the class applies such a test successfully the whole volume of the evidence may clearly be accepted as establishing the claim of the "smooth curve" to be the graphic representative of the formula.

§ 4. *Factorization* (Exs. V, VI).—With Ex. V begins the study of a number of topics that may be regarded as developments of the art of substitution. The first is *factorization.* As usually taught factorization is not a very elevating exercise. The beginner can hardly appreciate its value as an introduction to the study of "algebraic form". It can be justified to him only as a means of doing very simple multiplications and divisions in the head. But, since he has probably never seen any reason for doing these multiplications or divisions, factorization has, in effect, to be accepted as one more "rule"—a trick, amusing or depressing according to the

skill and temperament of the student. The optimist may, in addition, believe vaguely that it is a trick which some day may prove to have a use.

On the principles of this book, the usefulness of factorization should be made clear at the outset. Nothing is easier, for "identities" such as $ac \pm bc = (a \pm b)\, c$ and $a^2 - b^2 = (a + b)\,(a - b)$ have this obvious value: *they can be used to reduce the labour of arithmetical computation.* Here, then, is a clear reason for inviting the pupil to study them. There need be no fear—here or elsewhere—that the grossness of the utilitarian motive will destroy his sensitiveness to "algebraic form". On the contrary, in accordance with a universal psychological law, he will arrive all the sooner at a genuine interest in form if he sees how significant it may be from the point of view of economy of thought and labour.

The factorization of an algebraic expression is, then, to be taught, in the first instance, as a device which enables us to throw a formula into the shape most suitable for substitution. This principle decides both the range and the mode of treatment of the identities to be studied. They will be limited to the forms $ac \pm bc = (a \pm b)\, c$ and $a^2 - b^2 = (a + b)\,(a - b)$ together with forms readily derived from these. A kindred principle justifies the addition of the expansions of $(a \pm b)^2$ and $(a \pm b)^3$ to the young algebraist's armoury. All other identities are best postponed to a later period.

Detailed suggestions for teaching these identities are given in ch. VII. Here we need note only the following points. (*a*) The identities are introduced by instances in which their truth and their labour-saving virtue are obvious. This accounts (for example) for choosing the problem of *Exercises*, fig. 19, as the starting-point of ch. VII., B, instead of that of *Exercises*, fig. 24. (*b*) The process called "multiplication" is *subsequently* employed to show that the identities hold good universally (ch. I., § 5). (*c*) The formal elaboration of the identities is itself guided by the requirements of practical problems. This point is illustrated in Ex. V, **Nos. 5, 14,** and Ex. VI, **Nos. 6, 7, 24.** (*d*) It is an excellent plan to have copies of the figures of Exs. V and VI cut from cardboard or metal and to circulate them or use them with the class as a whole. Models of the solids of which these figures are cross-sections are still more useful, and to construct them in any suitable material is a valuable exercise.

The general principle followed in ch. VII. is to start with certain obvious geometrical relations and to base upon them an algebraic identity. This method is a reversal of the usual plan in which the identity is used to "prove" the truth of the geometrical propositions. Nevertheless it has historical justification. In the absence of a convenient numerical notation the Greeks were unable to develop an effective symbolic algebra (ch. I., § 5). A good deal of what we call their "geometry" was, as we have seen, really intended to supply this defect. Thus in Euclid, Bks. V and VII-X, the central interest is not in geometrical but in arithmetical analysis. Strictly speaking, lines and figures are there employed as the most convenient medium for expressing general arithmetical truths; in other words, these books contain largely a graphic algebra. Now, though the graphic presentation of a law or a relation is in the long run not so effective as its presentation in a formula (ch. IV., § 1), yet it has great merits, especially in the earlier stages of the subject. It presents the law in a less abstract medium than the symbolism, so that the untrained mind can more easily grasp it. This advantage of a graphic presentation is especially marked in cases like those of Exs. V and VI where the geometrical relations can be demonstrated by actually moving parts of a figure constructed of paper or cardboard. The teacher may be surprised to find how strikingly superior this method is to the mere contemplation of a blackboard drawing—superior as leading both to a much more ready discovery of the geometrical truth and a much more vivid appreciation of it. These psychological facts underlie the treatment of ch. VII. The algebraic presentation of the identity is based upon the graphic presentation, and each step in the manipulation of the symbols is simply a record of a corresponding manipulation of a tangible figure.

§ 5. *Square Root* (Ex. VII).—The calculation of square roots may be regarded as a process subsidiary to substitution. It enables the boy to find by a simple computation results that had to be found in Ex. IV by graphic interpolation. Ch. VIII. suggests a method of evaluation based on the principles of the preceding section. It is practically identical with that described about A.D. 365 by Theon of Alexandria.[1] It has two

[1] See an article and letters in the *School World* for January, February, and April, 1911.

great advantages. (i) It is thoroughly heuristic; i.e. the class may be expected to suggest spontaneously all the essential stages of the process. (ii) It is easily remembered or reconstructed if forgotten—a fact which illustrates the superior teaching value of what the psychologist calls "visual imagery".[1]

§ 6. *The Radical Notation. Surds* (Ex. VIII).—It is convenient at this point to consider the devices by which calculations involving the taking of square roots can be made with the least labour. Under this head we may limit our consideration to (i) the use of such a form as $\sqrt{12}$ as a symbol to be translated into a numerical value only at the end of the calculation; (ii) the obtaining of equivalences such as

$$\sqrt{12} = 2\sqrt{3};$$

(iii) the advantage of "rationalizing the denominator" in the case of such a fraction as $3/\sqrt{7}$. These operations should be taught and exemplified in connexion with Ex. VIII. The more complicated cases in which the denominator is a binomial surd are reserved to Ex. XXVI, E.

In dealing with this topic the teacher should remember that a "surd" such as $\sqrt{2}$ does not in elementary work stand for any definite number. It means a number which when squared will give as close an approximation to 2 as the nature of a particular problem demands.

If our present inquiry had a purely scientific instead of a practical aim we should find it necessary to go beyond this point and to find a definition of $\sqrt{2}$ that should distinguish it from any of the approximations which the arithmetical process of "finding the square root" can yield. The necessity for doing so arises from the connexion between number and geometrical magnitudes. Consider a right-angled triangle with two sides each of unit length. The length of the hypotenuse, AB, of such a triangle can be shown to be $\sqrt{2}$. For practical purposes this may be taken to mean that the length is 1·41 or 1·414 or 1·414235 . . . according to the degree of accuracy of measurement possible. Strictly speaking, however, each of these numbers represents a point situated

[1] Cf. *Report of L.C.C. Conference of Teachers*, January, 1911, discussion on "Memory".

at a different distance from A, and however far the process is continued it will never give the distance from A of the point B itself. Thus we must either abandon the idea that the distance AB can be exactly measured at all or else we must invent a special number to measure it—a number, that is, which has no place in the decimal scheme. Mathematicians have proposed different definitions to meet this need. That of Mr. Bertrand Russell[1] is the latest, simplest and most attractive. Such a symbol as 8 may be taken to represent two distinct numbers: (1) the ordinary "natural" or cardinal number eight; and also (2) the whole class of "rational" numbers, integral or fractional, which are less than eight. We can then define $\sqrt{2}$ as representing the class of rational numbers which when squared give rational numbers less than two. It is clear, on the one hand, that the "irrational number" thus defined is not a member of the ordinary decimal scheme, and, on the other hand, that it may be used quite unambiguously as the numerical label of the point B at the end of our hypotenuse.[2]

It is equally clear that this attempt at philosophical precision would be entirely out of place at the present stage of the course, and that the pupil should be taught to regard $\sqrt{2}$ as the symbol for a number which yields, when squared, a number differing from two by an unimportant amount.

§ 7. *The expansions of* $(a \pm b)^2$, $(a \pm b)^3$. *Approximation-formulæ* (Exs. IX-XI).—The square root process of ch. VIII. is based directly upon a study of fig. 17. In Ex. IX, **No. 2**, the pupil deduces from this figure the identity $(a + b)^2$, and in **No. 16** shows that it holds good universally. In the other examples the identity is used to reach an approximate solution of many problems in which an exact solution is either unnecessary or out of place because it would give a result too "fine" to be tested by measurement. The principle involved is very simple. If BX is $(1/n)$th of AB (fig. 17) the square Q (fig. 18) must be $(1/n^2)$th of the whole square BD. For example,

$$\text{if } BX = AB \times 0{\cdot}01, \quad Q = BD \times 0{\cdot}0001.$$

Thus if the calculation took account of only two decimal places the square Q could be neglected. This principle may be

[1] *Principles of Mathematics*, Vol. I, ch. XXXIII.

[2] The whole question raised here is discussed in *Exercises*, Part II, Ex. LXXII, and in ch. XXXIX. of this book.

simply illustrated by drawing several figures like fig. 16, the side AB being the same in each while ED is progressively smaller. The eye witnesses to the still more rapid progress towards insignificance of the area FD.

In Ex. IX, B, this principle is applied to the approximate evaluation of square roots. Thus in **No. 25** we have

$$\sqrt{18} = \sqrt{4^2 + 2}.$$

Taking the area of XY (fig. 17) as 16 the area of the strip of fig. 19 is 2, and the double base of RR′ is 8. If Q is ignored we can say that the area of R + R′ = 2 and the height BX = 2/8 = 0·25. Thus, ignoring Q, $\sqrt{18} = 4{\cdot}25$. In general, if p is small compared with a^2 we have

$$\sqrt{a^2 + p} \doteqdot a + \frac{p}{2a}$$

as the approximate value of the square root.

Ex. X illustrates the same principles in connexion with the identity $(a - b)^2 = a^2 - 2ab + b^2$. The only fresh point of interest lies in the manipulation by which fig. 17 can be made to exhibit the identity (**No. 1**). Putting

$$a = \text{AB and } b = \text{BX}$$

we have that ab is the rectangle composed of R + Q or that composed of R′ + Q. To reduce the square

$$\text{BD} = a^2 \quad \text{to} \quad \text{XY} = (a - b)^2$$

the following operations are therefore necessary. Take away R + Q, replace Q, take away R′ + Q. In symbols:—

$$\begin{aligned}(a - b)^2 &= a^2 - ab + b^2 - ab \\ &= a^2 - 2ab + b^2\end{aligned}$$

In Ex. XI the same principles are illustrated in connexion with the identities $(a \pm b)^3$. Some critics of authority have argued against the retention of these identities in elementary syllabuses, but the demonstration of the former by a model is so easy and attractive that there seems no sufficient reason for excluding it. There is another much weightier reason against exclusion. The natural field of application of the expansion of $(a + b)^3$ is to the properties of volumes. It plays the same part here as the expansion of $(a + b)^2$ plays in connexion with areas. At a time when it is generally recognized that elementary mathematics should give more and not less time to tridimensional problems it seems unwise

to dispense with an effective and simple instrument of spatial inquiry. Any valid objections to the identity in question are probably avoided in Ex. XI by limiting the examples almost entirely to problems on approximation.

The expansion of $(a + b)^3$ is most easily taught by means of the model[1] shown in fig. 7. The cube $(a + b)^3$ is to be built up from the cube a^3 by the following additions: (i) three slabs of dimensions $a \times a \times b$; (ii) three prisms measuring $a \times b \times b$ to fill up the spaces marked p, q, r; (iii) a cube, b^3, to fill up the space still left empty. Thus we have

$$(a + b)^3 = a^3 + 3a^2b + 3ab^2 + b^3$$

The additions are exhibited as a flat slab in fig. 8. (The letters refer to the problem described on p. 75.) It is evident that as $b/(a + b)$ decreases the cube b^3 soon becomes insignifi-

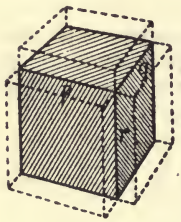


FIG. 7.

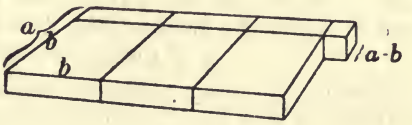


FIG. 8.

cant even when compared with $3ab^2$, the volume of the prisms. As the fraction continues to decrease, the row of prisms itself becomes insignificant compared with the three square slabs $3a^2b$. Thus if b is small enough we have :—

$$(a + b)^3 \doteqdot a^3 + 3a^2b$$

This result makes it easy to calculate a cube root approximately (**No. 10**). Let the number be written as $a^3 + p$; then $p = 3a^2b$ and $b = p/3a^2$. Thus $\sqrt[3]{a^3 + p} = a + p/3a^2$ approximately.

The same model may be used to exhibit the expansion of $(a - b)^3$, (**No. 8**), by operations similar to those used in connexion with $(a - b)^2$. It must be possible to remove together any one of the square blocks with the adjacent prisms and the small cube, and also to replace the prisms and cube separately.

[1] The model is most easily made of cardboard or stiff paper.

Call the whole cube a^3 and the depth of the parts to be removed b. Then a^3 can be turned into $(a - b)^3$ by operations performed in the following sequence :—

$$(a - b)^3 = a^3 - a^2b + ab^2 - a^2b + ab^2 - b^3 + ab^2 - a^2b$$
$$= a^3 - 3a^2b + 3ab^2 - b^3$$

The model may also be used to discover the identity $a^3 - b^3 = (a - b)(a^2 + ab + b^2)$ given as **No. 89** of Ex. XVI. Taking the larger cube of fig. 7 as a^3 and the smaller cube as b^3, the difference $a^3 - b^3$ may be exhibited in the form of the uniform slab of fig. 8. The thickness of this slab is $a - b$, its area $3ab + (a - b)^2 = a^2 + ab + b^2$. Hence its volume $= (a - b)(a^2 + ab + b^2)$.

Lastly it should be noted that Exs. IX-XI, important in themselves as illuminating the meaning and usefulness of identities which the pupil is apt to regard as barren and rather irritating abstractions, have a further importance as aiding the development of the notions upon which the later study of the calculus must be based. For this reason the teacher will do well to emphasize the question of the relative importance of the first, second and third powers of a small number and to illustrate it, when possible, by models. A series of models like fig. 7 with varying proportions of b to a is very useful for this purpose.

§ 8. *Fractions* (Exs. XII, XIII).—The proper significance of the process called simplifying an algebraic fraction has been pointed out in ch. I., § 5. Its aim is to predict by the manipulation of symbols the result that would be reached by simplifying an arithmetical expression containing fractions of a given type. Thus an "algebraic fraction" is, strictly speaking, no fraction at all. It is only a shorthand description of a certain class of actual (i.e. arithmetical) fractions. For instance $a/(a + 5)$ is a shorthand description of the class of fractions in which the denominator is greater by 5 than the numerator.

The discussions of ch. IX. place the study of algebraic fractions upon the same basis as that of factorization. It is "worth while" because by "simplifying" a fractional formula we can generally turn it into a form more suitable than the original for purposes of computation.

In Section I the range of fractions studied is limited to those with simple binomial denominators and monomial

numerators. Skill in manipulating such fractions suffices for the solution of most ordinary problems of real interest. On the other hand occasion is taken to complete the doctrine of approximations built up in Exs. IX-XI. Thus in Ex. XIII, **No. 15**, we have:—

$$1 + \frac{a}{1-a} = \frac{(1-a)+a}{1-a} \text{ and } a + \frac{a^2}{1-a} = \frac{(a-a^2)+a^2}{1-a}$$

$$= \frac{1}{1-a} \qquad = \frac{a}{1-a}$$

whence $$\frac{1}{1-a} = 1 + \frac{a}{1-a} \quad \ldots \qquad \text{(i)}$$

$$= 1 + a + \frac{a^2}{1-a} \quad \ldots \qquad \text{(ii)}$$

When a^2 is negligibly small this relation becomes

$$1/(1-a) = 1 + a.$$

Under the same conditions

$$1/(1+a) \doteq 1 - a$$

(**No. 16**). To obtain a "second approximation" (**No. 31**) we use the relation $a^2 + a^3/(1-a) = a^2/(1-a)$. Substituting for $a^2/(1-a)$ in (ii) we have:—

$$\frac{1}{1-a} = 1 + a + a^2 + \frac{a^3}{1-a} \quad \ldots \qquad \text{(iii)}$$

Thus to a second approximation $1/(1-a) \doteq 1 + a + a^2$. The process could evidently be continued indefinitely.

Fig. 9.

Fig. 10.

Fig. 11.

The examples of Ex. XIII, B, give a number of instances of the usefulness of these interesting approximations. Figs. 9-11 show a simple method of exhibiting to the eye the degree of approximation reached in a given case. Fig. 9 represents

the case of $1/(1 - a)$ when $a = \frac{1}{2}$. The whole strip measures the full value of the fraction, i.e. 2. The shaded part measures the value of the approximation $1 + a$, i.e. $1\frac{1}{2}$. The case when $a = \frac{1}{3}$ is dealt with similarly in fig. 11. Fig. 10 carries the case $a = \frac{1}{2}$ to a second approximation. The teacher will recognize in the whole treatment a useful preparation for the study of geometrical progressions (Ex. XXXVI).

§ 9. *Changing the Subject of a Formula.*—Exs. XIV, XV bring us to the important subject usually described as the solution of simple equations. The procedure advocated in ch. x. differs materially, both in spirit and method, from the current treatment of this central topic of elementary algebra. In history equations began as conundrums, and the school tradition has not lifted them to a much higher level of intellectual dignity. The pupil may become skilful in compelling "x" to reveal the value hidden in a symbolic statement of baffling complexity; he may become acute in threading the intricate mazes called "problems" which the ingenuity of the text-book writer has set in his path. Yet in the end he may still be only an expert solver of conundrums. He may have gained but an imperfect idea either of the practical or of the scientific importance of processes which he has learnt to handle for merely artificial purposes. We may well agree with Dr. Whitehead [1] that "one of the causes of the apparent triviality of much of elementary algebra is the preoccupation of the text-books with the solution of equations".

Nevertheless it would be bad tactics to ignore the pedagogical value of the conundrum. In all its varieties—from the riddle to the tragic "mystery"—it may be a powerful stimulus to intellectual activity. Thus in ch. x. the treatment is openly based upon the attraction of the conundrum, but this attraction is used to beguile the pupil into a study of processes which are immediately put to a more serious use. The nature of that use was explained in ch. I., § 5, where the practical value of the rules established in ch. x. was shown to consist in their power of leading us from old truths to new by an infallible mechanical process. By these rules any

[1] *Introduction to Mathematics* (Home University Library), p. 18.

variable which enters into a formula can be made the subject of the formula. For this reason the process may appropriately be called "changing the subject of a formula".[1]

Two points in the exposition of ch. x. should receive special attention. The first is that the rules for removing a number from one side of the formula to the other are not based upon any axioms of equality. This departure from traditional procedure needs but little justification. A boy is told that when 7 is added to a certain number the sum is 12, and at once states that the unnamed number was 5. It will not be pretended that he reaches this result by reflecting that "if equals be taken from equals the remainders are equal," nor that if he could not reach it unaided an appeal to that axiom would help him to conviction. Children can, in fact, solve such concrete riddles years before they can appreciate the abstract axiom.[2] Moreover the pupil will see with perfect clearness that the mode of solution of this problem—"take 7 from 12 and you have the other number"—is perfectly general; that is, that its validity as a process does not depend upon the specific numbers involved in it (ch. i., § 2).

These psychological considerations point the teacher to the plan followed in ch. x. The pupil first solves mentally a series of simple arithmetical problems. He is then led to *analyse* his solutions in symbols in order to make clear the principles underlying them. Finally he formulates the results of this analysis in "rules" of universal application (ch. i., § 2).

The mode of transition from solving a numerical riddle to changing the subject of a formula is the second point of importance. The reader will see that it is a particular application of the doctrine that symbols do not stand for numbers

[1] Since this phrase has already obtained a certain amount of currency, the author may be permitted to claim here the modest credit of its paternity. He believes that it was used for the first time in his lectures to teachers of mathematics in 1909. It was subsequently adopted in the Report on the Teaching of Algebra by the Committee of the Mathematical Association.

[2] Cf. the Board of Education's *Circular on the Teaching of Geometry and Graphic Algebra*. On the significance of axioms in algebra see the present writer's article in the *Mathematical Gazette*, for January, 1912.

but describe them (ch. I., § 3). There is, perhaps, no point at which the practical value of that doctrine is more clearly shown. The class whose teacher adopts the method given in ch. X., § 3, will find little difficulty in "literal equations" or in the much more interesting and important exercise of changing the subject.

Numerical results obtained as answers of the problems in Exs. XIV, XV should, of course, always be tested by substitution. In Ex. XIV, B, **Nos. 32-7,** the pupil is instructed to test also any formula derived from a given formula by changing the subject. Thus in **No. 32** he is, from the formula $W = b + mn$, to derive the formula $n = (W - b)/m$ by mechanical application of the rules of ch. X. He is then to think out afresh the problem of finding the number of marbles in a bag, given W, b and m, and to observe that he obtains the same formula as before. The object here is not to confirm the original answer so much as to convince the pupil of the labour-saving virtue of the rules. It is, in fact, another signal advantage of these exercises that they illustrate so clearly the "economy of thought" at which mathematics constantly aims. "Civilization advances by extending the number of important operations which we can perform without thinking about them."[1]

Finally it should be noted that the word "equation" is avoided throughout Section I. There seem good reasons for withholding the term until Section II, where, with the introduction of directed numbers, it becomes appropriate to use the typical form $f(x) = 0$ and to associate a new technical name with it (Ex. XXXVII).

§ 10. *Revision* (Ex. XVI).—Changing the subject of a formula completes the tale of the fundamental algebraic operations. Ex. XVI is, therefore, given to the revision and extension of the results gained during this first stage of the pupil's progress. The examples are classified into seven groups, A to G. A and B are rather harder examples of formulation and substitution and need no comment. In C the pupil is confronted with a number of the numerical puzzles and odd relationships which invariably tickle the interest of boys and girls both in and out of school, and is given the task of finding out how they "work". The skilful

[1] Whitehead, *Introduction*, p. 61. See also Mach, *Science of Mechanics, passim.*

teacher will find it possible, by a wise use of the gaiety appropriate to the theme, to make valuable use of these examples in exercising the analytical powers of his class. The method described in ch. VI., § 1, is the one generally to be followed: the pupil should work out a particular case of the puzzle and should then analyse his solution in symbols. On the other hand, if he is able to write down his analysis without previous consideration of a particular case he should, of course, be encouraged to do so.

Examples D give further exercise in graphic methods already studied in Ex. I. An account of all these methods will be found in ch. IV. **No. 60** is particularly useful because it introduces the third dimension. The teacher is recommended to take some pains, if necessary, to secure careful solutions of this example.

There are no novelties in E until **No. 87** is reached where the pupil is introduced to the method of "completing the square" which is afterwards to be used so frequently in the solution of quadratic equations and in other problems. The note before No. 87 gives sufficiently fully the argument which is intended to elucidate the process. It will be found a good plan to cut out large copies of the figures and to make members of the class actually carry out the "completion of the square" upon the blackboard. In **No. 88** the process is to be applied in order to calculate the length of the side of the square in given circumstances.

Examples F may be passed over as involving no new principles, and we may turn at once to the important argument developed in examples G, **Nos. 105-118.** From the standpoint of the usual classification these are examples of "quadratic equations". The teacher is, however, strongly recommended to resist the temptation to regard them as instances of a new "rule," but to treat them simply as cases in which the ordinary rules for changing the subject of a formula require supplementing by a little additional manipulation. The stages in the evolution of this manipulation are marked in Nos. 105, 109 and 110. In **No. 105** the first step is to change the subject of the formula to v^2; the second, to obtain a formula for v by simply taking the square root of the right-hand side. These steps are so obvious that the pupil will take them without prompting. **No. 109** requires a slight variation in the method; the necessary steps are

$$(s+a)^2 = \mathrm{A}$$
$$s+a = \sqrt{\mathrm{A}}$$
$$a = \sqrt{\mathrm{A}} - s$$

Finally, in **No. 110** the process of "completing the square" must be invoked in aid before the method of the last example can be applied. Thus we have

$$a^2 + 12a = 64$$
$$a^2 + 12a + 36 = 100$$
$$(a+6)^2 = 100$$
$$a+6 = 10$$
$$a = 4$$

It will be noticed that there is no indication in these solutions of the "ambiguity of sign" of a square root. The explanation is that as long as we are confined to non-directed numbers there is no such ambiguity, for the simple reason that the numbers have no signs. It will, indeed, be readily seen that there cannot be two answers to the question, What is the length of the side of a square of given area? and the suggestion that there are two possible values for the square root would in this case be only misleading. Nevertheless it is easy to find cases in which a problem of this type has two answers, though both are non-directed numbers. The note before **No. 118** explains the circumstances in which double answers are possible. It will be noted that the ambiguity which gives rise to them concerns, not the sign of a square root, but the order of the two terms of a binomial when the sign connecting them is a *minus*. It is extremely desirable to defer the consideration of the former type of ambiguity to the stage when the pupil is ready to study the "parabolic function" $ax^2 + bx + c$ in which the variables are assumed capable of positive and negative values, and to reserve the name "quadratic equation" for the problem represented typically by the symbolism

$$ax^2 + bx + c = 0$$

Problems of these kinds will come before us in Exs. LXII and LXIII.

CHAPTER VII.

FACTORIZATION.

A. Factorization of ac ± bc (ch. VI., § 4 ; Ex. V).[1]

§ 1. *The Identity* ac + bc = (a + b)c *discovered.* — The

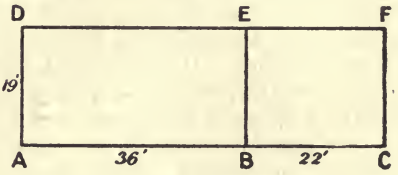


FIG. 12.

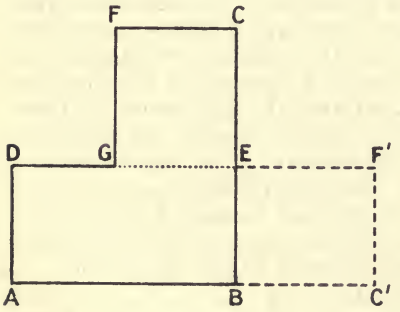


FIG. 13.

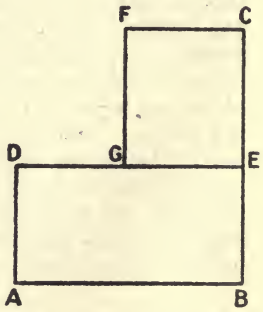


FIG. 14.

figure AF (fig. 12) is a plan of two rooms which have different lengths but the same breadth. Suppose AB = 36 feet, BC =

[1] A piece of cardboard shaped like fig. 14 should be prepared and should, before the lesson, be cut half-way through along the line GE so that CG may be readily detached and placed in the position of the dotted rectangle in fig. 13.

22 feet and AD = 19 feet. Then the joint area of the two rooms would be found by the calculation

$$A = 36 \times 19 + 22 \times 19$$

That is, it would be found by doing two multiplication sums and adding the results.

But we can calculate the joint area in a shorter way than this. Suppose the partition BE to be removed. Then instead of two rooms we shall have one room, 58 feet long, whose area must be the same as the sum of the areas of the original two rooms. (The partition may be supposed so thin that no allowance need be made for the area it stands on.) Thus the calculation is reduced to a single multiplication. The steps by which the reduction has been brought about can be written down as follows:—

$$\begin{aligned} A &= 36 \times 19 + 22 \times 19 \\ &= (36 + 22) \times 19 \\ &= 58 \times 19 \end{aligned}$$

The working of such a problem will always be shortened by this device, and the shortening may sometimes be very great. Thus if we wanted to know the total area of two rooms whose lengths were 63 feet and 37 feet respectively while they were both 29 feet wide, we should have:—

$$\begin{aligned} A &= 63 \times 29 + 37 \times 29 \\ &= (63 + 37) \times 29 \\ &= 100 \times 29 \\ &= 2900 \text{ square feet.} \end{aligned}$$

In this example the working is so much simplified that it can easily be done in the head.

The same plan could be used for shortening the work if we had to find the area of two rooms arranged as in fig. 14, or of a single room shaped like fig. 13. For since GE is of the same length as BE the rectangle FE can be supposed cut off at GE and placed in the position shown by the dotted rectangle. The area of the whole rectangle AF′ can then be calculated as before.

The rule we have just found can be used, of course, in calculations concerning *any* two rooms or other rectangular areas so long as they have one dimension in common. If a and b are used as symbols of the two unequal dimensions and

c as the symbol of the common dimension, the rule can be written thus :—

$$\begin{aligned} A &= ac + bc \\ &= (a + b)c \end{aligned}$$

[Ex. V, Nos. 1 to 5, may now be taken.]

§ 2. *The Identity* ac − bc = (a − b)c *discovered.*—Let us turn to a different kind of area-problem. The rectangle AD (fig. 15) represents the wall of a passage 32 feet long and 12 feet high. The passage is to be wainscoted up to EF, a distance of 4 feet 6 inches, and to be painted above that height. How many square feet are there to be covered with paint?

There are evidently two ways in which this calculation can be made. One way is to calculate the whole area AD, which

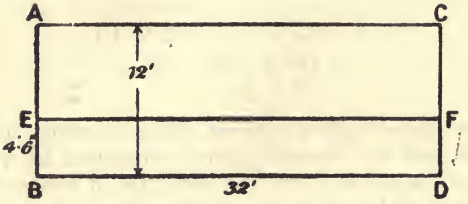


Fig. 15.

is 32 feet by 12 feet, and deduct the area of the wainscot ED which is 32 feet × 4½ feet. Thus :—

$$A = 32 \times 12 - 32 \times 4\tfrac{1}{2}$$

But it would obviously be quicker to calculate straight away the area of AF, whose length is also 32 feet and whose height is 7½ feet, that is 12 feet − 4½ feet. Setting down the steps by which the calculation is thus reduced to its simplest form, we have :—

$$\begin{aligned} A &= 32 \times 12 - 32 \times 4\tfrac{1}{2} \\ &= 32 \times (12 - 4\tfrac{1}{2}) \\ &= 32 \times 7\tfrac{1}{2} \end{aligned}$$

It is clear that this rule can be applied to find the area of any rectangle that can be considered as the difference of two rectangles which have one dimension in common. We have :—

$$\begin{aligned} A &= ca - cb \\ &= c(a - b) \end{aligned}$$

In the formulæ which we have written the letters have been placed in exactly the same position as the numbers to which they correspond in the calculations. This is generally speaking the best thing to do, since it shows clearly the connexion between the formula and the arithmetical working which the formula describes. But in the present problem there is no reason why we should take such pains to make our formula imitate the arithmetic, for we should not always write the arithmetic down in the same way. For example, the last calculation might on another occasion have been written—

$$\begin{aligned} A &= 12 \times 32 - 4\tfrac{1}{2} \times 32 \\ &= (12 - 4\tfrac{1}{2}) \times 32 \\ &= 7\tfrac{1}{2} \times 32 \end{aligned}$$

Corresponding to this the algebraic description of the working would be

$$\begin{aligned} A &= ac - bc \\ &= (a - b)c \end{aligned}$$

Thus the same calculation can sometimes be set down in two or more ways each of which can be described in a formula. Since, however, one of the things which a formula is meant to do is to serve as a convenient memorandum of a rule, there is an advantage in writing it always in the same way. It is usual, therefore, when the symbols can be written in different orders to choose some definite order such as the order of the alphabet. Thus we shall generally write $ac - bc$ rather than $ca - cb$, and $(a - b)c$ rather than $c(a - b)$. Sometimes, however, it may be convenient to reverse the usual order and, of course, it can never be actually wrong to do so.

§ 3. *When one of the Rectangles is a Square.*—Suppose in fig. 12 that BF was a square. Then the common dimension of the two rooms would in this case be the same as one of the two unequal dimensions—namely the one denoted by the symbol b. Thus the simplest way to represent the area of BF is by the symbol b^2, that is bb. It is unnecessary, therefore, to have another symbol, c, for the common width of the two rooms, since b will suffice. The rule for shortening the calculation now becomes :—

$$\begin{aligned} A &= ab + b^2 \\ &= (a + b)b \end{aligned}$$

Similarly if we take away a rectangle whose dimensions are

represented by the symbols a and b from a square the length of whose sides is also represented by a, we have

$$A = a^2 - ab$$
$$= a(a - b)$$

[Ex. V, Nos. 6-14, may now be taken.]

§ 4. *The Identities can be used in all Calculations.*—These rules are so useful in simplifying calculations of areas that we shall naturally inquire whether they cannot be used to simplify other calculations also. It will easily be seen that they can be so used.

Taking the first rule first, we note that it applies to calculations in which we have to find the sum of two products that possess one factor in common—for example $13 \times 7 + 13 \times 3$. Now whatever these numbers measure—or even if they are not meant to be measurements of anything at all—it is clear that 3 thirteens added to 7 thirteens must give 10 thirteens. That is

$$13 \times 7 + 13 \times 3 = 13 \times (7 + 3)$$

Similarly if we had to work out the calculation

$$13 \times 7 - 13 \times 3$$

it is plain that 3 thirteens taken from 7 thirteens will leave 4 thirteens, or that

$$13 \times 7 - 13 \times 3 = 13 \times (7 - 3)$$

Moreover we can obviously argue in the same way whatever numbers are substituted for the 13, 7 and 3, so long as one factor in the two products is the same. Thus using the symbol c to mean now, not necessarily a common *length*, but simply a number which is the common factor of two products, while a and b represent the unequal factors of the same products, we have

$$ac + bc = (a + b)c$$
$$\text{and } ac - bc = (a - b)c$$

For greater brevity these two results can be given together in the form

$$ac \pm bc = (a \pm b)c$$

which is to be read "*ac plus* or *minus bc*," etc.

The expressions which are linked here by the sign = simply represent different ways of carrying out the same calculation. For this reason they are often called *identities*.

[The remaining examples of Ex. V may now be worked.]

B. Factorization of $a^2 - b^2$ (ch. VI., § 4; Ex. VI).[1]

§ 1. *The Discovery of the Identity.*—Fig. 16 may be supposed to be the plan of a courtyard. The shape would be a complete square if there did not project into it a building which cuts off from the corner a smaller square, FD.

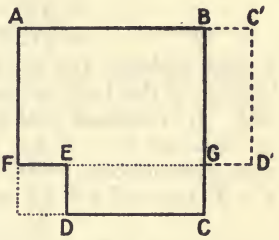


Fig. 16.

Let $a \equiv$ the length of the side of the larger square and $b \equiv$ the length of the side of the square FD; then the formula for the area of the courtyard is obviously

$$A = a^2 - b^2$$

Can we turn this into a form that would make calculation of the area easier? Our experience with the examples of the previous lesson suggests the inquiry whether it is possible to cut off the rectangle EC and place it so that with AG it forms one large rectangle. In order that this may be done either EG or ED must fit on to BG.

Let us first examine EG. Since $FG = a$ and $FE = b$, $EG = a - b$. But since $BC = a$ and $GC = ED = b$, $BG = a - b$ also. Thus EC can be cut off from the plan and placed in the position BD′.

We have now one rectangle AD′ whose area is the same as that of the plan. But $AC' = a + b$ (since BC′ is GC) and $C'D' = BG = a - b$

Hence
$$A = a^2 - b^2$$
$$= (a + b)(a - b)$$

Thus instead of squaring the two given lengths and subtracting one result from the other we can find the area in question by multiplying together the sum and difference of the given lengths. Since in the second method there is only one multiplication instead of two it will generally be an easier way of carrying out the calculation. Sometimes it will be a very much easier way. Thus if the side of the larger

[1] A large cardboard copy of ABCDEF (fig. 16) should be prepared beforehand and half cut through along the line EG, so that at the proper moment the rectangle EC may be detached and pinned on the blackboard in the position BD′.

square is 186 feet long and that of the smaller 86 feet long, we have :—

$$\begin{aligned} A &= (186)^2 - (86)^2 \\ &= (186 + 86)\ (186 - 86) \\ &= 272 \times 100 \\ &= 27200 \text{ square feet.} \end{aligned}$$

In this example the calculation is reduced to one that can be done in the head because one of the resulting factors is 100. It is obvious that there will also be great advantage in replacing $a^2 - b^2$ by $(a + b)\ (a - b)$ whenever a and b represent numbers which are nearly equal. For example, if $a = 97$ feet and $b = 92$ feet, we have

$$\begin{aligned} A &= (97)^2 - (92)^2 \\ &= 189 \times 5 \\ &= 945 \text{ square feet.} \end{aligned}$$

[Ex. VI, Nos. 1 to 10, may now be worked.]

§ 2. *The Identity can be used in any Calculation.*—We found in the last lesson that the expression $ac + bc$ may be replaced by $(a + b)c$ not only in area-formulæ but also in any other formula in which it appears. For this reason (it was said) $ac + bc = (a + b)c$ is called an *identity*. It is evidently important to find whether $a^2 - b^2$ can also be replaced by $(a + b)\ (a - b)$ in any formula.

As before we must examine cases in which the numbers are not intended as measurements of lengths or of anything else in particular. For example let a be supposed to stand for the number 9 and b for the number 6. Then

$$\begin{aligned} 9^2 - 6^2 &= 81 - 36 \\ &= 45 \end{aligned}$$

while

$$\begin{aligned} (9 + 6)\ (9 - 6) &= 15 \times 3 \\ &= 45 \end{aligned}$$

So that in this case it is true that

$$9^2 - 6^2 = (9 + 6)\ (9 - 6)$$

whatever the 9 and 6 measure.

Each member of the class should verify the rule in another case chosen by himself.

Let us next try to find out how this result comes about. The best way to do so is to start with $(9 + 6)\ (9 - 6)$ and try to show that the multiplication leads to the result $9^2 - 6^2$.

The expression $(9 + 6)\ (9 - 6)$ means, of course, that

$(9 + 6)$ or 15 is to be multiplied by $(9 - 6)$ or 3. We can calculate the result by taking 6 times $(9 + 6)$ from 9 times $(9 + 6)$, for the residue will obviously be 3 times $(9 + 6)$. That is

$$(9 + 6)(9 - 6) = (9 + 6) \times 9 - (9 + 6) \times 6$$

But $(9 + 6) \times 9$ (i.e. 9 fifteens) is evidently the same as 9 nines together with 9 sixes; and $(9 + 6) \times 6$ is the same as 6 nines and 6 sixes. We have, therefore, to take 6 nines and 6 sixes away from 9 nines and 9 sixes. Let us put this down in figures:—

$$\begin{aligned}(9 + 6)(9 - 6) &= 9 \times 9 + 6 \times 9 \\ &\quad - 9 \times 6 - 6 \times 6 \\ &= 9^2 - 6^2\end{aligned}$$

For $6 \times 9 = 9 \times 6$, so that the addition of 6×9 and the subtraction of 9×6 cancel one another.

We can deal similarly with any other difference of squares. For example:—

$$\begin{aligned}(13 + 5)(13 - 5) &= 13 \times 13 + 5 \times 13 \\ &\quad - 13 \times 5 - 5 \times 5 \\ &= 13^2 - 5^2\end{aligned}$$

Each member of the class should at this point analyse his own chosen example in the same way. Note that it is convenient so to arrange the figures that the cancelling products come one underneath the other.

It is evident that whatever numbers we take the same result will always follow. We can, therefore, describe the way in which the numbers behave by means of symbols. Putting a and b for the two numbers we shall have:—

$$\begin{aligned}(a + b)(a - b) &= a^2 + ba \\ &\quad - ab - b^2 \\ &= a^2 - b^2\end{aligned}$$

For whatever number ba represents ab must stand for the same number. Thus the addition of the number described by ba is always cancelled by the subtraction of the number described by ab, so that we are left with the difference between the squares of the two numbers.

[Ex. VI, Nos. 11 to 25, may now be worked.]

CHAPTER VIII.

SQUARE ROOT. SURDS.

A. The Calculation of Square Roots (ch. VI., § 5; Ex. VII).

§ 1. *The Simplest Case.*—If we know that the area of the square AC (fig. 17) is 16 square inches, then AB = 4 inches; if 25 square inches, AB = 5 inches, etc. But if we are informed that the area is 14·44 square inches, we cannot tell at a glance the length of AB. All that we know, at first, is that AB is between 3 and 4 inches. Mark off AX = 3

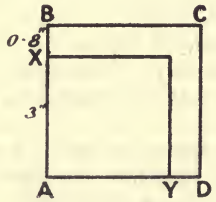


FIG. 17.

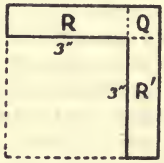


FIG. 18.

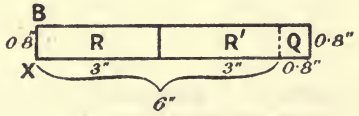


FIG. 19.

inches, draw the square XY of area 9 square inches, and suppose it cut out from the square AC. The residue (fig. 18) can be divided into two equal rectangles, R and R', and a square, Q, having a total area of 14·44 − 9 = 5·44 square inches. This residue can be re-arranged as in fig. 19 to form a long rectangle. The problem is to find XB, the height of the rectangle. If we knew the base of the rectangle we should

divide it into 5·44, the area of the whole strip, and so obtain the height exactly. We know that the base of

$$R + R' = 3 \times 2 = 6,$$

so that 6 divided into 5·44 will give us the height very roughly. The quotient is 0·9, but since this must be too big we will suppose that the height is really 0·8. On this assumption the base of the rectangle would be 6·8 inches and the height 0·8 inch. The area would be

$$6{\cdot}8 \times 0{\cdot}8 = 5{\cdot}44 \text{ square inches.}$$

But this is exactly the required area. Therefore XB = 0·8 and $\sqrt{14{\cdot}44} = 3{\cdot}8$. The various steps are best set down in the form :—

$$\begin{array}{r|l} & 14{\cdot}44\ (3 + 0{\cdot}8 \\ & \underline{9} \\ 6{\cdot}8 & 5{\cdot}44 \\ & \underline{5{\cdot}44} \\ & \end{array}$$

$$\therefore \sqrt{14{\cdot}44} = 3{\cdot}8$$

In a similar manner the square roots of 32·49 (5·7) ; 51·84 (7·2) ; 68·89 (8·3) should be reasoned out.

§ 2. *Non-terminating Roots.*—Suppose the area of the square to be 40 square inches. What is the length of AB? As before mark off AX = 6 inches, and remove the square XY so as to leave a residue R + R′ + Q = 40 − 36 = 4 square inches. Since the base of R + R′ = 6 × 2 = 12, the height XB is approximately 4/12, i.e. about 0·3 inch. Taking it as exactly 0·3 we have that the area of the strip

$$R + R' + Q = 12{\cdot}3 \times 0{\cdot}3 = 3{\cdot}69 \text{ square inches.}$$

This area is less than the required area by

$$4 - 3{\cdot}69 = 0{\cdot}31 \text{ square inch.}$$

The state of affairs we have reached is represented by fig. 20. We have removed from the whole area of 40 square inches the area of a square whose side is 6·3 inches, and we have a residue $r + q + r'$ whose area is 0·31 square inch. This residue can in turn be arranged as in fig. 19, except that the three sections of the rectangle must be labelled r, r' and q. We know that the base of $r + r'$ is 6·3 × 2 = 12·6 inches, so that the height of the

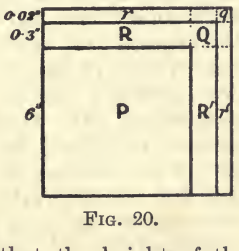


FIG. 20.

rectangle is rather less than 0·31/12·6, i.e. about 0·02 inch. This value would give for the whole rectangle $r + r' + q$ the area $12{\cdot}62 \times 0{\cdot}02 = 0{\cdot}2524$ square inch, leaving a deficiency of $0{\cdot}31 - 0{\cdot}25 = 0{\cdot}06$ square inch. That is, the area of a square $6 + 0{\cdot}3 + 0{\cdot}02 = 6{\cdot}32$ inches in the side would be only 0·06 square inch less than 40 square inches. We could carry the approximation still farther in exactly the same way, but as it is already so close that the error cannot be exhibited in the figure the labour would be superfluous. The only further step we need take is to find whether a square of 6·32 inches, or one of 6·33 inches in the side would give the nearer approximation to an area of 40 square inches. The residue 0·06 divided by $6{\cdot}32 \times 2$ gives a number less than 0·005. It is clear, therefore, that 6·32 inches is the value to be adopted. The working will be set down thus:—

	40	(6 + 0·3 + 0·02 +
	36	
12·3	4·00	
	3·69	
12·62	0·3100	
	0·2524	
12·64	0·0576	

$\therefore \sqrt{40} \doteq 6{\cdot}32$

§ 3. *Numbers greater than* 100 *or less than* 1.—Suppose our plans all have to be made on a scale of 1 foot to 10 feet. Every linear dimension would be 10 times as small as in reality, every area 100 times as small. Thus in the various cases considered the areas would really be 1444 square feet, 3249 square feet, 5184 square feet, 6889 square feet, and 4000 square feet, while the sides of the squares would be 38 feet, 57 feet, 72 feet, 83 feet, and 63·2 feet. This consideration gives us a rule for finding the square root of a number greater than 100. Divide the number by 100, find the square root and multiply it by 10. (This is of course only a question of the position of the decimal point.) If it is greater than 100×100, divide it by this number, take the square root and multiply the latter by 10×10, i.e. 100.

Similarly if the number is a decimal less than unity

multiply it by 100 or 100 × 100, etc., as the case requires, find the square root and divide by 10, or by 10 × 10, etc.

[Ex. VII may now be worked.]

B. The Radical Form. Surds (ch. VI., § 6; Ex. VIII).

§ 1. *Relations between Square Roots.*—We have seen (A, § 3) that if we know the square root of a number it is unnecessary to calculate the square root of a number 100 times as large or as small. The root of the second number is obtained by multiplying or dividing the root of the former by 10. This is an instance of a principle that is often useful. Suppose that we know that $\sqrt{40} = 6{\cdot}32$ approximately and want to find the square root of 360. Then we have:—

$$\begin{aligned} 360 &= 9 \times 40 \\ &= 3 \times 3 \times 6{\cdot}32 \times 6{\cdot}32 \\ &= (3 \times 6{\cdot}32) \times (3 \times 6{\cdot}32) \\ \text{whence } \sqrt{360} &= 3 \times 6{\cdot}32 \\ &= 18{\cdot}96 \text{ approximately} \\ \text{Similarly, since } 1{\cdot}6 &= 40/25 \\ \sqrt{1{\cdot}6} &= 6{\cdot}32/5 \\ &= 1{\cdot}26 \text{ approximately} \end{aligned}$$

We conclude that when a number is the product or the quotient of two numbers whose roots are known its root is the product or the quotient of their roots.

§ 2. *Use of the Radical Notation.*—While there are some numbers which have an exact square root—like 16, 178929 and 13·1044 of which the roots are 4, 423 and 3·62 respectively—the great majority of numbers really have no square root at all. These are called "surds". Thus it is evident from A, § 2, that the process of finding the square root of 40 would never come to an end. In other words, there is no decimal fraction which when squared yields exactly 40. All we can say is that by prolonging the process of A, § 2, we can get numbers which, when squared, come constantly nearer to 40. Thus the first stage of the process consists in observing that $7^2 > 40 > 6^2$. Since the length of the strip of fig. 19 is between 12 and 13 its height is between 4/13 and 4/12, that is, between 0·307 . . . and 0·333 . . . , or, say, between 0·3 and 0·4. We have, then, $(6{\cdot}4)^2 > 40 > (6{\cdot}3)^2$. By examining the successive strips $r + r' + q$, etc., we obtain the following series

of pairs of numbers such that the square of one is greater and the square of the other less than 40. The one whose square is nearer to 40 is in heavy type.

$7^2 = 49$	$> 40 >$	$\mathbf{36} = 6^2$
$(6{\cdot}4)^2 = 40{\cdot}96$	$> 40 >$	$\mathbf{39{\cdot}69} = (6{\cdot}3)^2$
$(6{\cdot}33)^2 = 40{\cdot}0689$	$> 40 >$	$\mathbf{39{\cdot}9424} = (6{\cdot}32)^2$
$(6{\cdot}325)^2 = \mathbf{40{\cdot}005625}$	$> 40 >$	$39{\cdot}992976 = (6{\cdot}324)^2$
$(6{\cdot}3246)^2 = \mathbf{40{\cdot}00056516}$	$> 40 >$	$39{\cdot}99930025 = (6{\cdot}3245)^2$
etc.		etc.

Although 40 has really no square root it is customary to say that its square root is one of the numbers 6, 6·3, 6·32, 6·325, 6·3246, etc., according to the degree of exactness required. Thus $\sqrt{40}$ occurring in a calculation must be taken to mean one of this series. It is generally convenient not to decide until the end of the calculation *which* of the numbers is to be taken. For this reason the expression $\sqrt{40}$ is often treated in calculations as if it were a symbol like a, the value of which is to be filled in after factorizing, etc. The following is a good example of the advantage of this procedure:—

§ 3. *Rationalizing the Denominator.*—In the formula of Ex. IV, No. 31, let D = 40. Then we have $P = 18/\sqrt{40}$. Now if at this point we substitute for $\sqrt{40}$ one of the series obtained in § 2 we shall have to carry out a decimal division. But by treating $\sqrt{40}$ as a symbol it is possible to avoid this inconvenient operation. We proceed thus:—

$$P = \frac{18}{\sqrt{D}}$$

$$= \frac{18}{\sqrt{40}}$$

$$= \frac{18}{\sqrt{40}} \times \frac{\sqrt{40}}{\sqrt{40}} \quad . \quad . \quad . \quad \text{(i)}$$

$$= \frac{18\sqrt{40}}{40}$$

$$= \frac{9\sqrt{40}}{20}$$

$$= \frac{9 \times 6{\cdot}32}{20} \quad . \quad . \quad . \quad . \quad \text{(ii)}$$

$$= 2{\cdot}844 \text{ lb.}$$

In (i) the denominator is "rationalized" by multiplying the fraction by $\sqrt{40}/\sqrt{40}$—an operation which leaves its value unchanged. The substitution of a value of $\sqrt{40}$ occurs only at the end of the calculation in (ii). Note that if we had known $\sqrt{10}$ instead of $\sqrt{40}$ our method would have been $P = 18/2\sqrt{10} = 9/\sqrt{10} = 9\sqrt{10}/10$.

[Ex. VIII may now be worked.]

CHAPTER IX.

FRACTIONS.

A. Fractions with Monomial Denominators (ch. I., § 5; ch. VI., § 8; Ex. XII).

§ 1. *Reduction of Formulæ containing Fractions.*—A cistern is being fed by a ball-tap which would fill it in forty-eight minutes. At the same time water is being drawn out by a pipe which would empty it in fifty-three minutes. If it is empty to begin with, in how many minutes will it be full? For the solution we have:—

Fraction of cistern filled in 1 minute

$$= \frac{1}{48} - \frac{1}{53} \quad . \quad . \quad \text{(i)}$$

$$= \frac{53 - 48}{48 \times 53} \quad . \quad . \quad \text{(ii)}$$

$$\therefore \text{time taken} = \frac{48 \times 53}{53 - 48} \quad . \quad . \quad \text{(iii)}$$

It is obvious that we should have followed the same method whatever numbers had been given. The process can, therefore, be described in words and expressed in symbols. Put $t_1 \equiv$ "the number of minutes in which the tap would fill the cistern," and use t_2 and T similarly as the other symbols required. Then we have the following "shorthand" description of steps (i), (ii), (iii):—

$$\frac{1}{T} = \frac{1}{t_1} - \frac{1}{t_2} \quad . \quad . \quad . \quad . \quad \text{(i)}'$$

$$= \frac{t_2 - t_1}{t_1 t_2} \quad . \quad . \quad . \quad . \quad \text{(ii)}'$$

$$T = \frac{t_1 t_2}{t_2 - t_1} \quad . \quad \text{(iii)}'$$

Now it is clear that formula (iii)′ is more convenient for calculating T than formula (i)′, for its use enables us to dispense with two steps of arithmetic. Moreover, if the arithmetical calculation goes through steps like (i), (ii) and (iii) the symbolic description of those steps must always proceed by corresponding steps like (i)′, (ii)′ and (iii)′. Thus if we are given a formula like (i)′ we can always work through a stage corresponding to (ii)′ to a formula corresponding to (iii)′ which is in the form most suitable for calculation.

§ 2. *Rules for Manipulating Formulæ.*—There is no need of special rules—like the rule $a^2 - b^2 = (a + b)(a - b)$—for carrying out these changes in a formula. The one sufficient rule is, obviously, the following: Whenever a formula contains symbolism which describes a fractional expression it may be simplified for purposes of calculation by manipulating the symbols in exactly the same ways as the figures would be manipulated in the arithmetic which the symbolism describes. Thus, suppose we are given a formula $1/R = a/b + c/d$ and are asked to simplify it for calculation. The fact that we do not know what words the symbols stand for does not hinder us in the least, for we know that the right-hand symbolism describes the sum of two fractions; a and c being the description of the numerators, b and d of the denominators. The steps involved in the arithmetical calculation must, therefore, be those described by the steps:—

$$\frac{1}{R} = \frac{a}{b} + \frac{c}{d}$$

$$= \frac{ad + bc}{bd}$$

$$R = \frac{bd}{ad + bc}$$

§ 3. *The Least Common Denominator.*—If the denominators of two fractions contain a common factor that factor occurs only once, not twice, in the denominator used to express the sum or difference of the fractions. For example $\frac{7}{8} + \frac{5}{12} = \frac{3 \times 7 + 2 \times 5}{4 \times 2 \times 3}$. If we know that the denominators of the fractions described in a formula have a common factor it is easy to arrange the statement in symbols so as to show

the effect of this circumstance upon the calculation. Thus, putting p for the common factor we should have :—

$$\frac{a}{pb} + \frac{c}{pd} = \frac{ad + bc}{pbd}$$

[Ex. XII may now be worked. No. 1 may be taken orally.]

B. Fractions with Binomial Denominators (ch. I., § 5; ch. VI., § 8; Ex. XIII).

§ 1. *Algebraic Fractions.*—It would be tiresome always to say that "$\frac{a}{b}$ represents or describes a fraction," that "a describes the numerator," and that "b describes the denominator". For the sake of brevity we can say that $\frac{a}{b}$ *is* a fraction, that a *is* its numerator and b its denominator. This is merely repeating what we did when we agreed to use the form "let the length of the room be l feet," instead of the longer form, "let l be the symbol for the length of the room in feet". In both cases for the sake of brevity and convenience we speak of symbols which stand for the descriptions of numbers just as if they were themselves numbers. When we want to make it clear that we are speaking of these "fractions" made up of symbols and not of the real fractions of arithmetic, we will call them "algebraic fractions". It should be noticed that the number described by the algebraic fraction a/b may not even be an arithmetical fraction at all, but a whole number. This will be so, for instance, if, in a given case, we have $a = 12$, $b = 3$.

§ 2. *One Binomial Denominator.*—In previous examples of algebraic fractions the numerators and denominators have been either single symbols such as c or products of single symbols[1] such as pq or p^2. There is, however, no reason why the denominator of a fraction should not sometimes be more conveniently expressed as the sum or difference of two symbols.

For an example suppose that the pipe of A, § 1, takes five minutes longer to empty the cistern than the ball-tap takes to

[1] Note that pq is really a symbol for the product of two numbers represented by p and q. It is called a product of the symbols p and q for brevity.

fill it. In these circumstances if we represent the latter time by t the former time is best represented by $t + 5$. The original formula now becomes :—

$$\frac{1}{T} = \frac{1}{t} - \frac{1}{t + 5}$$

This fractional expression can be simplified in exactly the same way as before. For although $t + 5$ is a more complicated symbol, yet, after all, it represents a single number just as the symbol t_2 does in A, § 1. The only difference between what we must do now and what we did then is that when we manipulate $t + 5$ we should enclose it in brackets. This practice will remind us that $t + 5$ really represents a single number which, if we were doing arithmetic, would be moved *as a whole* from one place in the expression to another. We have, then,

$$\frac{1}{T} = \frac{1}{t} - \frac{1}{t + 5}$$

$$= \frac{(t + 5) - t}{t(t + 5)}$$

But the numerator of the new fraction can obviously be simplified. For if we take t away from $t + 5$, whatever t may represent, the residue must always be 5. Thus, the next line of the working will be

$$= \frac{5}{t(t + 5)}$$

while, finally :—

$$T = \frac{t(t + 5)}{5}$$

In general, if the difference of time, which in this case is five minutes, is represented by the symbol d, the work will read as follows :—

$$\frac{1}{T} = \frac{1}{t} - \frac{1}{t + d}$$

$$= \frac{(t + d) - t}{t(t + d)}$$

$$= \frac{d}{t(t + d)}$$

$$\therefore T = \frac{t(t + d)}{d}$$

Instead of taking t to represent the time in which the tap fills the cistern let it represent the time taken by the pipe to empty the cistern. Then the former time must now be represented by $t - 5$, and the formula becomes :—

$$\frac{1}{T} = \frac{1}{t-5} - \frac{1}{t}$$

$$= \frac{t - (t-5)}{t(t-5)}$$

$$= \frac{5}{t(t-5)}$$

$$\therefore T = \frac{t(t-5)}{5}$$

We have seen in Ex. VI, Note to No. 15, how to simplify the numerator of the fraction in the second line. We have to take $t - 5$ away from t, that is, to take from t the whole of t except 5. Whatever number t represents the residue must, of course, be 5. We can, if you prefer it, argue in another way. We have to take away $t - 5$, that is, a number 5 less than t. If we take away t we shall have taken away 5 too much and must give back 5 to make the account square. But if we take t from t we shall have nothing left, so that when we restore the 5 this number will be the total. By both arguments, then, $t - (t - 5) = 5$.

The conclusion must hold good for all numbers. Thus whatever numbers a and b represent we have :—

$$\frac{1}{a-b} - \frac{1}{a} = \frac{a - (a-b)}{a(a-b)}$$

$$= \frac{b}{a(a-b)}$$

[Ex. XIII, Nos. 1-3 and 7-33, can now be worked.]

§ 3. *Two Binomial Denominators.*—We may now go on to consider other kinds of fractional expressions that are likely to occur in formulæ. We shall then be prepared to deal with them when they present themselves in actual problems.

It is very likely, for example, that we shall sometimes meet with a fractional expression slightly more complicated than the last, such as $\frac{1}{a-7} + \frac{1}{a+3}$. Let us begin by finding an expression more suitable for calculation by which this could

be replaced. Remembering that $a - 7$ and $a + 3$ are really symbols for single numbers such as 20 and 30 (i.e. $27 - 7$ and $27 + 3$) we have :—

$$\frac{1}{a-7} + \frac{1}{a+3} = \frac{(a+3)+(a-7)}{(a-7)(a+3)}$$
$$= \frac{2a-4}{(a-7)(a+3)}$$

There is no difficulty about simplifying the numerator. We have to add to the number $a + 3$ a number which is 7 less than a. This can be done by adding a and taking away 7. Thus

$$(a+3)+(a-7) = a+3+a-7$$
$$= 2a-4$$

Consider next the expression in which the fractions of the last expression are connected by a *minus* instead of by a *plus* sign :—

$$\frac{1}{a-7} - \frac{1}{a+3} = \frac{(a+3)-(a-7)}{(a-7)(a+3)}$$
$$= \frac{10}{(a-7)(a+3)}$$

To find the simplified numerator we have to take the number $a - 7$ from the number $a + 3$. We can do this in two ways.

We can take $a - 7$ from a and then add 3. But if we take $a - 7$ from a the result is, as we know, 7. Thus, the numerator would be 10. This argument written in symbols would be

$$\{a-(a-7)\}+3 = 7+3$$
$$= 10$$

Note that we put a pair of curled brackets round the expression $a - (a - 7)$ to show that we are thinking of it as representing a single number.

Or we can say that instead of taking away $a - 7$, that is, a number 7 less than a, we shall reach the same result by taking away a and adding 7. Expressed in symbols this argument reads :—

$$(a+3)-(a-7) = a+3-a+7$$
$$= 10$$

By similar arguments the numerator $(a - 3) - (a - 7)$ could be simplified in either of the following ways :—

$$\begin{aligned}(a-3)-(a-7) &= \{a-(a-7)\}-3\\ &= 7-3\\ &= 4\end{aligned}$$

$$\begin{aligned}(a-3)-(a-7) &= a-3-a+7\\ &= 7-3\\ &= 4\end{aligned}$$

What should we do if the numerator were $13-(a-7)$? In this case the first argument would be rather round-about. There is no a from which to take $a-7$. It we want one we must add it to the expression and at the end take it away again. Thus we may write:—

$$\begin{aligned}13-(a-7) &= a+13-(a-7)-a\\ &= \{a-(a-7)\}+13-a\\ &= 7+13-a\\ &= 20-a\end{aligned}$$

The other argument is, this time, much simpler. We take away a from 13 and then add 7. Thus

$$\begin{aligned}13-(a-7) &= 13-a+7\\ &= 7+13-a\\ &= 20-a\end{aligned}$$

The only difficulty about the second argument is that a might be greater than 13, for example 15. We could not take 15 away from 13 and then add 7. The first argument, however, always holds good. It shows us that even if a is greater than 13 it is still correct to write

$$13-(a-7)=13-a+7$$

The reason why it is correct is, of course, that to take away a and then add 7 ought to produce the same result as adding 7 and then taking away a. If a is greater than 13 we cannot do the first thing but we can do the second.

§ 4. *Numerators not Unity.*—The fractions whose combinations we have studied in this lesson have all had the same numerator, namely unity. There will, of course, be many occasions when this is not the case. Another "cistern problem" will serve to exemplify this fact. Suppose the cistern to be fed by four taps, each capable of filling it in $t-5$ minutes, while water is being drawn off by three pipes, each of which would empty it in $t+2$ minutes. Then the time taken to fill the cistern when all the taps and pipes are at work is given by the formula:—

$$\frac{1}{T} = \frac{4}{t-5} - \frac{3}{t+2}$$
$$= \frac{4(t+2) - 3(t-5)}{(t-5)(t+2)}$$

Before we proceed further we must know how to simplify the numerator $4(t+2) - 3(t-5)$.

To begin with, $4(t+2)$ means 4 times a number made up of t and 2. This product must, of course, be the same as four ts together with four 2s, or $4t + 8$. Similarly the product represented by $3(t-5)$ must be the same as the number $3t - 15$. Thus we have

$$4(t+2) - 3(t-5) = (4t+8) - (3t-15)$$

The further simplification of this expression can be carried out by either of the two methods which we have used before.

Thus we may argue:—

$$(4t+8) - (3t-15) = \{(3t - (3t-15)\} + t + 8$$
$$= 15 + t + 8$$
$$= t + 23$$

Or we may produce the result of taking away the number $(3t - 15)$ by first taking away $3t$ and then adding 15:—

$$(4t+8) - (3t-15) = 4t + 8 - 3t + 15$$
$$= t + 23$$

As before, the second is the better practical method. It has a simple rule: remove the brackets and change the sign connecting the terms within them.

The simplification can now be completed. We have:—

$$\frac{1}{T} = \frac{4}{t-5} - \frac{3}{t+2}$$
$$= \frac{4(t+2) - 3(t-5)}{(t-5)(t+2)}$$
$$= \frac{t+23}{(t-5)(t+2)}$$
$$\therefore T = \frac{(t-5)(t+2)}{t+23}$$

[Ex. XIII, Nos. 4-6 and 34-6, may now be taken.]

CHAPTER X.

CHANGING THE SUBJECT OF A FORMULA.[1]

§ 1. "*Think of a Number*" *Problems.*—"I am thinking of a number. I multiply it by 2 and subtract 7 from the product. The result is 11. What is the number?" The whole class see at once that the number is 9. One member is then asked to describe the steps by which he reached this conclusion. "If after taking away 7 from twice the number 11 is left, twice the number must be 18. If twice the number is 18, the number itself must be 9." These steps are now, at the dictation of the class, to be recorded on the blackboard[2]:—

$$\text{(i)}\quad \begin{aligned} 2n - 7 &= 11 \\ 2n &= 11 + 7 \\ &= 18 \\ n &= \frac{18}{2} \\ &= 9 \end{aligned}$$

"Think of a number" questions of each of the following forms should be asked and the modes of solution analysed in the same way. That is, the answer should be obtained mentally, the steps by which it is reached described verbally, and the description transcribed in symbols. In addition to the former example the blackboard should now exhibit the following or similar matter:—

$$\text{(ii)}\quad \begin{aligned} 3n + 6 &= 30 \\ 3n &= 30 - 6 \\ &= 24 \\ n &= 24/3 \\ &= 8 \end{aligned} \qquad \text{(iii)}\quad \begin{aligned} 3(n - 4) &= 18 \\ n - 4 &= 18/3 \\ &= 6 \\ n &= 6 + 4 \\ &= 10 \end{aligned}$$

[1] See ch. I., § 5; ch. VI., § 9; Exs. XIV, XV.

[2] It is vital to note that what is written on the blackboard is *merely a shorthand transcription of the answer just given.* It is not "doing it by algebra instead of by arithmetic".

$$\begin{aligned} \text{(iv)}\ \tfrac{1}{2}(n + 7) &= 5 \\ n + 7 &= 5 \times 2 \\ &= 10 \\ n &= 10 - 7 \\ &= 3 \end{aligned}$$

§ 2. *Analysis and Generalization of the Method.*—A study of these four examples shows that they are solved by the use of two rules: (A) If the expression on the left-hand side of the sign " = " is written as a sum (or difference) the number to be added (or subtracted) is removed and is subtracted from (or added to) the number on the right of the sign. (B) If the expression on the left is written as a product (or a quotient) the numerical factor (or divisor) is removed and becomes a divisor (or a multiplier) of the number on the right. Rule A is exemplified in the second line of (i) and (ii) and the fourth line of (iii) and (iv); Rule B in the fourth line of (i) and (ii) and the second line of (iii) and (iv).

Care must be taken to apply each rule at the proper point. Thus in (iii) it would be wrong to begin by adding 4 to 18, because $3(n - 4)$ is written as a product not as a difference. We could, however, easily turn it into a difference thus:—[1]

$$\begin{aligned} 3(n - 4) &= 18 \\ 3n - 12 &= 18 \\ 3n &= 30 \\ n &= 10 \end{aligned}$$

reaching the same result as before.

Again, it would be wrong to begin in (ii) by dividing the 30 by 3 (i.e. to write $n + 6 = 10$) because $3n + 6$ is written as a sum not as a product. We can, however, turn it into a product thus:—[1]

$$\begin{aligned} 3n + 6 &= 30 \\ 3(n + 2) &= 30 \\ n + 2 &= 10 \\ n &= 8 \end{aligned}$$

Thus, although these problems can often be solved mentally in more than one way, Rules A and B hold good whichever way is taken.

Lastly it is clear that the rules hold good in all cases. Thus in (i) it is not because the number to be subtracted on

[1] Note that the algebra is, once more, only a transcription of the verbal explanation.

the left is 7 that it may be removed and added on the right. We should begin solving the problem in the same way whatever number stood in the place of the 7. Similarly the division in the fourth line is not performed because the multiplier on the left is 2; we should deal in the same way with any number standing in the place of the 2.

Now suppose the following question is asked (Ex. XIV, No. 1): "I am thinking of a number. I multiply it by 3·6 and add 14·7. The result is 23·18. What is the number?" This problem is too hard for most of us to do in our heads, but that fact need not prevent us from solving it. We can write the statement down in symbols and can find the number required simply by applying to it Rules A and B.

[Ex. XIV, Nos. 1-6, may be taken here.]

§ 3. *Description of Method in Symbols.*—Since a "Think of a number" problem could be solved in the same way even if you changed all the numbers in the statement, it is possible to describe in symbols the methods followed in solving each kind. Consider, for example, problems like (i) in § 1. Let $a \equiv$ "the number multiplying n," $b \equiv$ "the number to be subtracted on the left," $c \equiv$ "the number on the right". Then the method of solution can be described, line by line, as follows:—

$$\begin{aligned} an - b &= c \\ an &= c + b \\ n &= \frac{c+b}{a} \end{aligned}$$

The actual working of No. 6 and the description of the method are here set side by side.

$$\begin{aligned} \frac{n}{4\cdot4} + 7\cdot35 &= 13\cdot6 \\ \frac{n}{4\cdot4} &= 6\cdot25 \\ n &= 6\cdot25 \times 4\cdot4 \\ &= 27\cdot5 \end{aligned} \qquad \begin{aligned} \frac{n}{a} + b &= c \\ \frac{n}{a} &= c - b \\ n &= a(c-b) \end{aligned}$$

[Ex. XIV, Nos. 7-31, may now be taken.]

§ 4. *Changing the Subject of a Formula.*—In Ex. III, No. 2 (ii) we found a formula for the salary of a clerk after so many years of service. Suppose now that we are asked for a formula for the number of years before his salary will reach a given amount. We can, of course, obtain the new formula

in the same way as we obtained the old one, by thinking out the rule and then writing it down in symbols. But by means of Rules A and B of § 2 we can obtain it in a more convenient way directly from the formula $S = S_o + ni$.

To begin with, the statement

$$S = S_o + ni \quad . \quad . \quad . \quad . \quad (i)$$

may also be written

$$S_o + ni = S \quad . \quad . \quad . \quad . \quad (ii)$$

for (i) says, "The salary after n years is obtained by adding ni to the original salary," while (ii) says, "Add ni to the original salary and you will obtain the salary after n years". Thus the two statements differ only in the order of the words. But in the formula which we are seeking S_o, i and S will be *descriptions of known numbers*—just as a, b, c, etc., were in § 3—while n is here, just as it was there, a symbol for a number which is not given but has to be calculated. It follows that we may apply Rules A and B to i and S_o in (ii):—

$$ni = S - S_o \quad . \quad . \quad . \quad . \quad (iii)$$
$$n = (S - S_o)/i \quad . \quad . \quad . \quad (iv)$$

in order to obtain a formula, (iv), with n as subject instead of S. This operation may be called *changing the subject of the formula*. It is easy to satisfy oneself that the new formula is true.

§ 5. *The Meaning of the Symbols need not be known.*—We have, then, the following rules for changing the subject of a formula. First manipulate the formula so that the new subject appears on the left. (This manipulation will often consist in interchanging the two sides of the formula—a process which may be called Rule C.) Next apply Rules A and B to the other symbols until they are all on the right-hand side of the sign of equality. The result is the desired formula.

The most interesting thing about this process is that you do not have to know what a formula means in order to change its subject. If the first formula is true and Rules A, B and C are performed correctly, the resulting formula must also be true. Take as examples the unexplained formulæ $P = \frac{a + bQ}{n}$ and $P = \frac{a}{bQ - t}$ and change the subject to Q in each case. For the first we have:—

$$P = (a + bQ)/n$$
$$(a + bQ)/n = P \quad \text{Rule C}$$
$$a + bQ = nP \quad \text{Rule B}$$
$$bQ = nP - a \quad \text{Rule A}$$
$$Q = (nP - a)/b \quad \text{Rule B}$$

For the second we have:—

$$P = a/(bQ - t)$$
$$(bQ - t)P = a \quad \text{Rule B}$$
$$bQ - t = a/P \quad \text{Rule B}$$
$$bQ = a/P + t \quad \text{Rule A}$$
$$Q = \frac{a/P + t}{b} \quad \text{Rule B}$$

[Ex. XIV, B, and Ex. XV may now be taken.]

CHAPTER XI.

THE PROGRAMME OF SECTION I (EXS. XVII-XXVI).

§ 1. *Direct and Inverse Proportion.*—The general course of the argument running through Exs. XVII-XXVI has been described in ch. v., § 2. Ex. XVII introduces the idea of "functionality," which is further developed in Exs. XXII-XXV. Strictly speaking, the use of a formula always implies this idea; for a formula always exhibits the value of one variable as depending upon the values of other variables. The aim of these exercises is to make the implicit idea explicit. In previous exercises the pupil has been chiefly interested in the *matter* of a formula—that is, in the information which it gives about the subject. He is now to see that formulæ which deal with a most heterogeneous collection of subjects may yet have precisely the same *form;* and this text is to be illustrated by a detailed study of a few specific functional relations.

Two features of the exposition are of special importance. In the first place the more abstract notion of algebraic form is approached by way of the more concrete graphical form. For example, the straight line and the hyperbola which are respectively the expressions of direct and inverse proportion are literally "forms". The student who has stored them in his memory has ever at hand a means of keeping hold of ideas which are apt to evade the grip of an algebraic expression. Their remembered shapes exhibit in a flash to the mental eye the typical features of each relation. For these reasons emphasis is laid in this part of the subject upon the correspondence between certain graphic and certain algebraic forms [cf. ch. IV., § 8].

In the second place those familiar denizens of the algebraic page—x and y—now come into view for the first time.

Letters have hitherto been the shorthand representatives of verbal descriptions of specific or concrete variables. For this reason they have always been chosen so as to suggest the variables to which they refer. The best way to record the fact that the weight of a piece cut from a sheet of cardboard is proportional to its area is to use the form $W = kA$. But if we wish to forget the specificity of the variables and to attend only to their form of connexion it is well to use letters which will not suggest any particular variables but shall mean only "variables in connexion with one another". For this purpose x, y and z will be reserved. It follows that, as our students come to deal more and more with variables in general instead of specific variables, xs and ys will the more abound. The teacher is strongly advised not to throw away the great advantage of this special use of x and y by using them as symbols for unknown values of a concrete variable.

The forms of relation (or "functions")[1] considered in Exs. XVII and XXII-XXV are (with one exception) limited to direct and inverse proportion. Ex. XVII deals with $y = kx$, Ex. XXIII with $y = k/x$—each function being approached by way of its graphic symbol. In Ex. XXIV the study of direct and inverse proportion is extended to the functions $y = kx^2$ and $y = k\sqrt{x}$, in the one connexion, and to the functions $y = k/x^2$ and $y = k/\sqrt{x}$ in the other. Attention should be given to the way in which the relation between the primitive and the derived form of proportion is brought out. Each form of direct proportion is in the first instance referred to the straight line through the origin; each form of inverse proportion to the hyperbola. The graphs of the more complex forms are then derived from the straight line or hyperbola by a simple method of transformation. In this way the pupil is taught to realize the unity underlying the various forms in which direct proportion and inverse proportion may be exhibited. Ex. XXV is given to a simple study of mixed forms, such as $z = kxy$, of such common occurrence that they could not be omitted even in the first stage.

The exception referred to above is the "linear relation" which grows so naturally out of the relation of direct proportion that its study could hardly be excluded. It appears in Ex. XVIII, Nos. 19-24.

[1] The *term* "function" is reserved for Section II.

§ 2. *The Trigonometrical Ratios.*—The reasons for including in the algebra course the fundamental notions of trigonometry have been given in ch. II., § 3. Methods of imparting them are suggested in ch. XIII. It will be seen that the subject is here treated as a department of the doctrine of direct proportion. The study of the sine, cosine and tangent as functions which have a "field" of their own apart from their relations to triangles is postponed to Part II, Section VII. At first the treatment is confined to these three fundamental functions; the secant, cosecant and cotangent are probably not worth the confusion they introduce.

All three notions are taught in connexion with practical problems, in which they appear as factors which when multiplied into the length of one side of a right-angled triangle give the length of another. The average pupil finds these definitions of tangent, sine and cosine much easier to understand and to apply than their definitions as ratios. The tangent is taught first (ch. XIII., A) since the practical problems in which it can be applied are the most obvious and simple. The sine and cosine are taught (ch. XIII., B) in connexion with navigation problems for two reasons besides the intrinsic attractiveness of such problems. First, the calculation of easting and northing from a given course and distance exhibits the sine and cosine in the rôle in which they play their most characteristic part in all branches of pure and applied mathematics—namely as factors for determining the "projections," "components" or "resolved parts" of some directed magnitude. Next, in navigation problems the right-angled triangle is in varied positions, so that the pupil learns from the outset to think of the sine and cosine as connecting the hypotenuse with the sides opposite and adjacent to a given angle. Taught in this way these notions do not contract the haziness which persistently clings round them when they are defined in terms of the "perpendicular" and "base" of a triangle.[1]

In connexion with the aim of illustrating the special use

[1] In order to eliminate a fruitful source of confusion the Greek letters α, β, γ will be used as symbols for the *measure* of an angle in degrees. The Roman letters A, B, C, etc., P, Q, R, etc., will be used (as in geometry) merely to name the angle. Thus if in fig. 32, p. 128, the angle P contains 34° we have $\alpha = 34°$. (Note that we do not write $\alpha°$.) Later in the work the symbols θ, ϕ will be used for the measure of an angle in radians.

of the sine and cosine the idea of a vector is introduced (Ex. XIX, Note to No. 18). The reader is no doubt familiar with this notion. In fig. 21, which illustrates Ex. XIX, **No. 20,** the vectors b and a represent two movements carrying a point from A to B. The vector c represents the single movement which would have carried the point from the same origin to the same terminus. Then b and a are component vectors, c their resultant. It is important to note that the angle between b and a is the external angle at C. It is the change of direction which one would make at B in walking

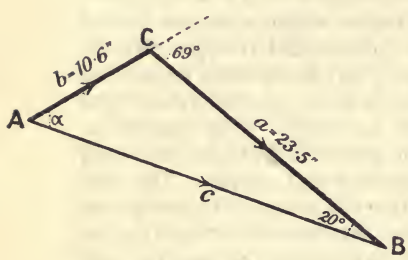


FIG. 21.

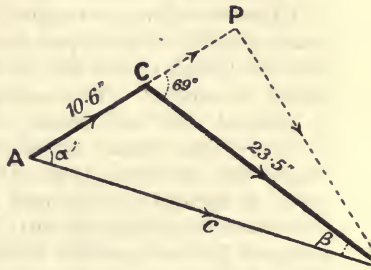


FIG. 22.

along the route ACB. The angle A = 69° − 20° = 49°. Hence

$$c = b \cos 49° + a \cos 20°.$$

By the aid of the vector notion all cases[1] of the "solution of triangles" can easily be solved without the aid of special formulæ. Thus if, in fig. 21, the information that ABC = 20° had been withheld (**No. 22**), the length of c could still have been calculated by the method shown in fig. 22. Suppose the point to travel from A to B by the rectangular route APB instead of by ACB. Then we have

AP = 10·6 + 23·5 cos 69° = 19, and PB = 23·5 sin 69°
= 22 (nearly).

But tan α = PB/AP = 22/19 = 1·16 = tan 49°.

Hence α = 49° and β = 20°,

while c can be calculated by two or three methods already studied.

[1] Except that of determining the three angles from the three sides.

Fig. 23 shows the solution of **No. 28:**

PB = 800 sin 53° = 639; a = PB/sin 78° = 639/sin 78° = 653.

Hence b = 800 cos 53° + 653 cos 78°.

Ch. XIV., A, continues the subject of navigation problems in order to exhibit the application of the trigonometrical functions to the simple geometry of the sphere. The neglect of this geometry in schools where globes and atlases are in constant use is little short of a scandal. Ex. XX is the beginning of an attempt continued throughout the book to give clear though strictly elementary ideas upon a subject which—apart from its special claims upon the interest of a seafaring people—offers the best possible opportunities for cultivating the knowledge of tridimensional space.

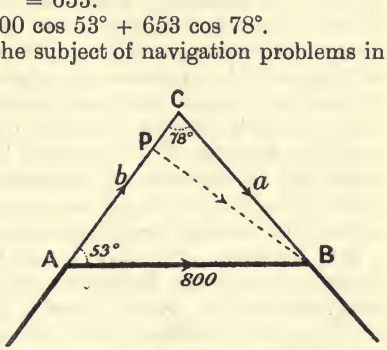


FIG. 23.

In ch. XIV., B, and Ex. XXI, the pupil is made acquainted with the fundamental relations between the three trigonometrical ratios of the same angle. It will be noted that they are taught not as abstract propositions but as practical tests of the accuracy of the ratios supposed to be determined by measurement. Any cultivation of "identities" beyond this point would be quite out of harmony with the general treatment.

§ 3. *The Combining of Formulæ* (Exercise XXII).—Between the earlier and the later exercises on proportionality an exercise dealing with "simple equations of two unknowns" is interpolated. The traditional treatment which takes this topic immediately or shortly after equations of one unknown simply because two follows after one is based upon the infertile principle of "logical" arrangement. In ch. XV. an attempt is made to find a more natural and fruitful mode of attack. There are three outstanding uses of systems of equations. The first is to determine the constants in a relation, such as $y = a + bx$, the form of which is already

known. The second is to determine whether any pair (or triad) of values of the variables will satisfy each of a given set of relations; and, if so, what is the pair or triad. The third use is elimination—the process of "editing" a set of symbolic statements so as to derive from them a statement in which all they have to say about the connexion between certain variables is included while all the irrelevant things they say about other variables are deleted. These three topics form the subjects of ch. xv. and Ex. XXII. It should be noted that the discussion is in each case based upon operations which the pupil has learnt in earlier exercises. The aim is to systematize these operations into technical methods.

§ 4. *Revision.*—Ex. XXVI and last of Section I is given to general revision. Revision of this kind is of great importance, especially in order to secure technical facility. But it is still more important that the pupil shall constantly be called upon to apply his knowledge outside the formal mathematics lesson. The making and manipulating of formulæ in physics—including change of the subject—statistical work and field-surveying in geography offer opportunities for the kind of application which makes mathematical knowledge at once a reality and a delight. A simple plan of co-ordination between the school departments will provide many such opportunities to the profit of all concerned and will secure the unity of method which is the essence of fruitful co-operation.

The bulk of the examples are arranged in six "test papers," printed in two blocks—A,B,C, and F,G,H. An important feature consists in the problems in which the pupil is called to solve simple problems of solid geometry by means of trigonometrical formulæ. Simple models constructed in paper or thin card are of great assistance in building up the power of dealing intelligently with problems of this kind.

Between the two sets of test papers are two groups of examples whose aim is to extend rather than to revise the earlier work. The group headed "E" deals with the manipulation of surds, and may be regarded as an appendix to Ex. VIII. Its main purpose is to familiarize the pupil with the process of "rationalizing the denominator" of a surd fraction when the denominator is a binomial. The group headed "D" is composed of examples of a more novel character. These aim at introducing certain simple methods which statisticians use in summarizing the information to be derived

from a given group of data. A little consideration will show that any such method involves the use of at least two numbers or "co-ordinates". Take as a simple example the attempts of a marksman to hit a certain point on a target. If we suppose that his shots have left visible marks upon the target it is clear that they may, in general, be regarded as constituting a swarm or constellation of points. With regard to this swarm we may ask two questions: (i) What is its general position on the target? and (ii) What is the degree of closeness or "scatter" of the individual shots? If by an astonishing fortuity the marks happened to lie in a perfect circle it would be easy to give a definite numerical answer to both those questions: the "general position" of the swarm would be defined by the position of the centre of the circle, the degree of "scatter" by its radius. The problem of the statistician is to find similarly definite replies in less simple cases. The examples before us illustrate two methods, either of which may appropriately be used when the data are non-directed numbers. The method most suitable for directed numbers (the method of "root-mean-square deviation") is studied in Ex. LVII. Both these incursions into the province of statistical science are to be regarded as preliminary to the formal treatment of that important and typically modern subject in the last section of the whole work.

In both of the methods here to be considered the "general position" of the group of data is regarded as determined by the position of the "median," or middle term of the series. There are two alternative ways of measuring the degree of dispersion or "scatter". The first is to find by calculation the arithmetical mean or average of the differences between the various data and their median. This number is called the "mean deviation" of the group of data. The other method is to divide the series into four compartments, each containing one quarter of the whole number of data arranged in order, and to find the semi-distance between the "quartiles" or points of division which lie on either side of the median and include half the series between them. It is obvious that this "quartile deviation" or "semi-interquartile range" indicates in a general way the degree of condensation or dispersion of the data about the median.

Of these two methods the latter has the advantage of requiring practically no calculation. It is also directly ap-

plicable when the data are represented graphically instead of numerically. **Nos. 13-20** illustrate this important point in connexion with the problem of evaluating the evidence afforded by a series of practical measurements. The way in which the application is made was described in ch. IV., § 7.

CHAPTER XII.

DIRECT PROPORTION.

(Ch. IV., § 6; ch. XI., § 1; Ex. XVII.)[1]

§ 1. *Direct Proportion: its Graphic Symbol.*—The idea of direct proportion is familiar from arithmetic. We say, for example, that the cost of a length of stair carpet is directly proportional to the number of yards bought; meaning that if on one occasion we buy 2, 3, 4 . . . times, or $\frac{1}{2}$, $\frac{1}{3}$, $\frac{2}{3}$, $\frac{4}{7}$, . . . as much carpet as on another occasion the cost will be 2, 3, 4 . . . times or $\frac{1}{2}$, $\frac{1}{3}$, $\frac{2}{3}$, $\frac{4}{7}$ . . . as much as before.

Suppose a man who sells stair-carpets at different prices to need unexpectedly a "ready reckoner". It could be made most expeditiously upon the following plan. On squared paper graduate a base-line to represent the number of yards of carpet sold (fig. 24), and a vertical line to represent the cost. Above the graduation 20 insert dots marking the price of 20 yards of each kind of carpet. We may suppose that there are four kinds, priced respectively at 2s. 3d., 3s. 1½d., 3s. 9d. and 4s. 6d. a yard. (To use the prices of 1 yard would be to risk inaccuracy in drawing; the cost of 20 yards is easily calculated mentally.) Through each dot and the origin draw the straight lines OA, OB, etc. These lines constitute the required "ready reckoner". For example, the cost of 48 yards of the cheapest carpet is given by the point P and is £5 8s.

§ 2. *Geometrical Proof.*—It is easy to show why the ready reckoner "works". Suppose OQ to be divided into any number of equal parts of which Oq_1 is the first, and p_1q_2, p_2q_3, etc., are equal to the others. Then all the triangles Oq_1p_1, $p_1q_2p_2$, etc., are congruent. It follows that the line

[1] The teacher should also consult the chapter on Proportion in Abbott's *Teaching of Arithmetic*, published in this series.

rises from O to P by equal steps, p_1q_1, p_2q_2, p_3q_3, etc. Thus if there are n steps, PQ $= p_1q_1 \times n$. Suppose q_1 to be taken at the point representing 1 yard. Then there would be 48 steps and PQ $= p_1q_1 \times 48$. Thus if p_1q_1 represents correctly the price of 1 yard, PQ will represent correctly the price of

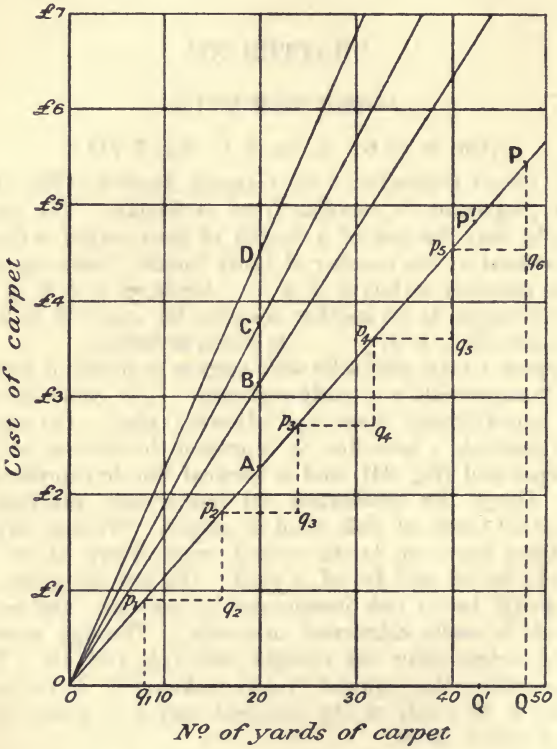


Fig. 24.

48 yards. But there would be 20 equal steps from O to A, so that, since the vertical at A represents correctly the price of 20 yards, p_1q_1 *does* represent correctly the price of 1 yard.

We are certain, then, that the points on the line OAP give correctly the cost of any exact number of yards. It is easy

to make sure that the reckoner will also "work" for fractions of a yard—for example, that the point P′ gives correctly the price of 42 yards 2 feet, or $42\frac{2}{3}$ yards = £$1\,2\,8$. For this result the line can be supposed to rise from O to P′ by 128 equal steps, of which the first should represent the cost of $\frac{1}{3}$ yard. But since 60 of these equal steps will carry the line to A, the first step does represent correctly the cost of $\frac{1}{3}$ yard. Thus P′Q′ represents correctly the cost of $42\frac{2}{3}$ yards. It is obvious that the argument could be repeated with any other fractional number of yards.

§ 3. *Correspondence between Graphs and Formulæ.*—The cost in pounds (C) of a length of L yards of the cheapest carpet is given by the formula

$$C = 0{\cdot}1125L$$

Since by the straight line OA the same problems can be solved as by this formula, the line may be called *the graph of the formula* C = 0·1125L. Similarly the line OD is the graph of the formula C = 0·225L, etc. In general, if p be the price of a single yard of carpet, then the formula C = pL will give the cost of a given number of yards, and to this formula will correspond some straight line through the origin—the slope of the line depending on the value of p.

§ 4. *Generalization of Formula.*—It is clear that we should have reached results like the foregoing if we had started with any other example of direct proportion. That is, no matter what the things are of which one is directly proportional to the other, the graph of the formula connecting them would always be a straight line through the origin. Such a straight line may, therefore, be regarded as an universal symbol of direct proportion.

It is convenient to have a formula which can also be used as an "universal symbol" of direct proportion. For this purpose we must first replace the letters C and L in the formula C = pL by letters that shall not suggest any definite "variables". The letters x and y are very suitable for this use. It is usual to let x represent the "independent variable"—that is, the variable whose values are supposed to be given. The letter y will then represent the "dependent variable"—that is, the variable which would be the subject of the corresponding formula. In accordance with the rule followed in graph-drawing (ch. IV., § 3), x, the unnamed independent variable, will always be measured along the horizontal axis.

For this reason this axis is often called *the axis of* x or the x-axis. The vertical axis is then called *the axis of* y or the y-axis.[1]

Lastly p must be replaced by a letter that will not suggest any definite example of proportion. The letter k will generally be used. It will suggest that in the formula corresponding to any straight line through the origin each value of y is obtained by multiplying the proper value of x by a *constant* number. With given scales of graduation for x and y, the value of k depends on the slope of the graph, being greater as the latter is steeper.

[Ex. XVII may now be worked.]

[1] The distance of a point P from the axis of x is called *the ordinate* of P; its distance from the axis of y *the abscissa* of P.

CHAPTER XIII.

THE TRIGONOMETRICAL RATIOS (I).

A. The Tangent of an Angle (ch. XI., § 2; Ex. XVIII).

§ 1. "*Height Problems*" *by Calculation.*—Standing 150 feet from the wall of a building I note that the angle of elevation of its top is 22°. How high is the top of the wall above my eye-level? This problem could, of course, be answered by means of a drawing (fig. 25). Draw OA representing 150 feet to any scale you please, set off OB so that the angle AOB = 22°. The perpendicular AB read off on the scale used in drawing OA gives the height of the wall. It will be found to be 60 feet.

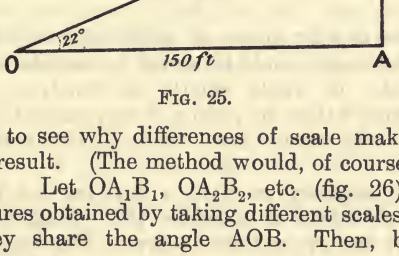


FIG. 25.

It is easy to see why differences of scale make no difference to the result. (The method would, of course, be useless if they did.) Let OA_1B_1, OA_2B_2, etc. (fig. 26), represent different figures obtained by taking different scales and placed so that they share the angle AOB. Then, by the last lesson we have:—

$$A_1B_1 = k \,.\, OA_1,\ A_2B_2 = k \,.\, OA_2, \text{ etc.},$$

k being a constant number. If, therefore, in any one drawing the perpendicular is (say) $\frac{2}{5}$ of the base, it will be so in all possible drawings representing the same data.

Now the value of k depends, as we have seen, upon the angle at O. When this angle is 22° it appears that $k = \frac{2}{5}$ or 0·4. Supposing this result to be accurate it would be possible to solve any other similar problem involving an angle of 22° by calculation and so to avoid the tedium and risk of drawing. For example: I stand 80 yards from the point immediately under the spire of a church and the weathercock has an elevation of 22°. How high is it above my eye-level? Suppose OA (fig. 25) *on any scale* to represent 80 yards; then AB represents the height of the spire. But we have

$$\begin{aligned} AB &= OA \times \tfrac{2}{5} \\ &= 80 \times \tfrac{2}{5} \\ &= 32 \text{ yards.} \end{aligned}$$

§ 2. *Extension to other Angles. The Table of Tangents.*— It is easily seen that to each angle there corresponds a

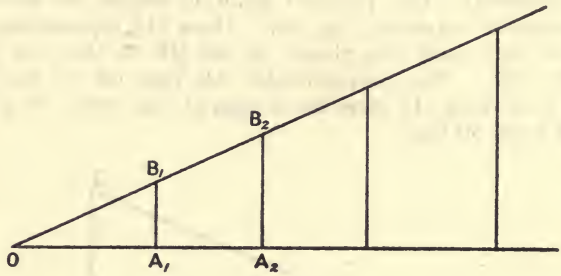


FIG. 26.

definite value of k by means of which all "height problems" involving that angle could be solved by calculation instead of measurement. It would clearly be worth while to make a table of these values by means of very careful drawings and to preserve it for future use. Such a table is called *a table of tangents.*

It is not necessary to make a drawing for every degree. It will be sufficient to measure the "tangents" (i.e. the values of k) for every tenth degree. The tangent for any other angle can be determined from these by a graph. The work should be divided among the class. Each member should draw a right-angled triangle containing one of the angles in the following table. To ensure accuracy the triangle should be large, but no attempt should be made

to ensure uniformity. It should be pointed out that each triangle supplies materials for calculating two tangents. For example, the person who starts out to determine the tangent of 40° can also calculate from his figure the tangent of 50°. The conclusion is drawn that the tangent of an angle is the reciprocal of the tangent of its complement. By this plan at least four independent measurements of each tangent will be obtained. The results should be collected upon the blackboard, averaged, and set out in the following table. The tangents of 0° and 90° must be discussed separately. In the case of the latter it must be shown that as the angle increases the tangent also increases in such a way that by bringing the angle near enough to 90° we can make the tangent as large as we please. Thus there is no number so large that we cannot find an angle whose tangent is larger still. Moreover, the tangents continue to grow larger with the angle until the latter becomes a right angle. Nevertheless, there is no tangent corresponding to the right angle itself, for at the moment when the angle becomes a right angle the figure ceases to be a triangle. (A triangle whose base is zero would be an absurdity.) To indicate (i) that there is no tangent corresponding to 90° and (ii) that the tangent increases without end as the angle approaches 90° it is usual to make use of the symbolism

$$\tan 90° = \infty .$$

This statement is usually read "the tangent of 90° is infinity". The pupil must remember that this does not mean that 90° has a tangent like the other angles, but that the words are a convenient brief way of expressing what was said above.

Angle	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
Tangent	0·00	0·18	0·36	0·58	0·84	1·19	1·73	2·75	5·67	∞

From these data the graph of fig. 27 should now be drawn. Individuals should be asked to read from their graphs the tangents of given angles. The accuracy of the results should be challenged by comparison with those of other individuals

and with the numbers in the Table of Tangents (*Exercises*, p. 107).

[Ex. XVIII may now be worked.]

C. The Sine and Cosine. Vectors (ch. XI., § 2; Ex. XIX).

§ 1. *The Navigator's Problem.*—The master of a ship which leaves one port for another, perhaps at a great distance, must be able at any moment to mark his position upon his *chart* or map of the seas. He can do this very exactly from time to time by observations of the sun from which he can calcu-

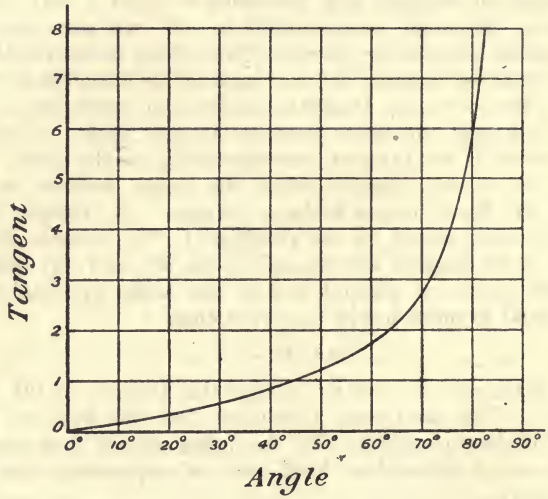


FIG. 27.

late his latitude and longitude. But such observations can be made only in favourable circumstances—e.g. in clear weather. They must be supplemented by other calculations which will fix the position of the ship at times between the observations on the sun.

These calculations—which the sailor calls his "dead reckoning"—are made as follows. When he has left a known point he can fix his position at a later moment if he knows two things: the direction in which he has sailed and the distance the ship has travelled in that direction. The former

is called the *course* and is determined by the compass. On steamers, at any rate, the course is usually reckoned by the number of degrees between the direction in which the ship is being steered and the north and south line. Thus if the ship is being steered W.S.W. the course is "$67\frac{1}{2}°$ W. of S." The *distance run* on a given course is determined by the *log*. In modern ships this is an instrument which can be compared with a cyclometer, since it records at any moment how far the ship has gone. (A brief description of the patent log may

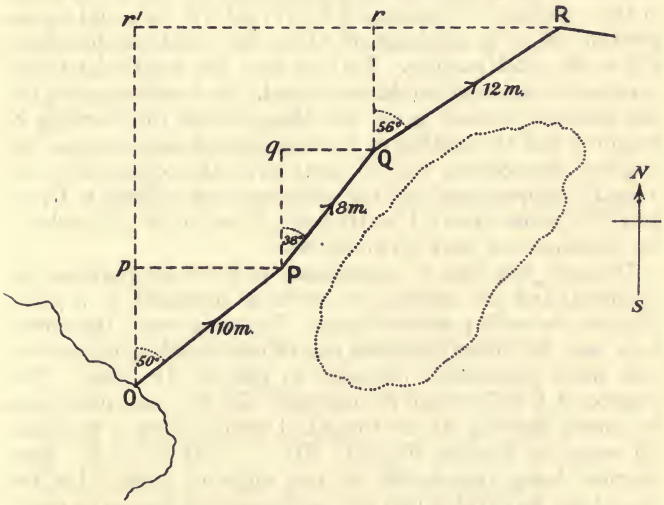


FIG. 28.

well be given, prefaced by an account of the older method of "heaving the log". The term "ten knots" which means "ten sea-miles an hour" receives its explanation here.) Whenever the direction of the ship is changed the new course is recorded in the "log book" together with the distance run upon the last course.

§ 2. *Graphic Solution.*—Knowing the distance run on each course the sailor could now fix his position on the chart. Let O (fig. 28) be the port of departure. Let the ship begin by steaming 10 miles on a course 50° E. of N. This movement can be represented by the line OP drawn to scale

and at an angle of 50° with the north and south line Op. The lines PQ, QR may similarly represent movements of the ship in which it runs 38° E. of N., 8 miles, and 56° E. of N., 12 miles, successively. The point R now gives the position of the ship according to the "dead reckoning". But there is a more convenient method than this. The movement of 10 miles along OP will carry the ship a certain distance to the north of O and also a certain distance to the east. The same thing is true of the other movements. If we knew the amount of the "northings" represented by Op, Pq, Qr, and of the "eastings" represented by pP, qQ, rR we could fix the position of R by marking off Or' = the total northing and r'R = the total easting. In this way the trouble and risk involved in drawing would be avoided. It is evident after the last chapter (*a*) that, given the distance run, the northing or southing and the easting or westing depend entirely upon the angle of the course; and (*b*) that, given the course, they are directly proportional to the distance run. Thus if Op is (say) 6·4 miles when OP is 10 miles, it would be 12·8 miles if the distance run were 20 miles, etc.

To apply this idea it is necessary to know the northing (or southing) and the easting (or westing) produced by a given distance for each possible course. Knowing these, the northings, etc., for other distances can be calculated by proportion. The most convenient distance to take is 10 miles. The method of A, § 2, should be followed. Let the class determine by careful drawing the northing and easting when a ship sails 10 miles on courses 10°, 20°, 30°, . . . 80° E. of N.—each member being responsible for one angle or more. Let the class then be divided into two sections—one section to graph the eastings (fig. 29), the other the northings (fig. 30).

§ 3. *Sine and Cosine.*—From these graphs, taken together, we can read off the northing (or southing) and the easting (or westing) corresponding to a run of 10 miles on any given course. For example if a ship runs 37° W. of S. she will be carried 8 miles south and 6 miles west of the starting point.

Suppose that, the course being the same, the distance run was 17 miles, we should argue that each mile run along that course carries the ship 0·8 miles south and 0·6 miles west. Thus for a run of 17 miles:—

Southing = 17 × 0·8 = 13·6 miles and Westing = 17 × 0·6 = 10·2 miles

The numbers 0·8 and 0·6 and the numbers corresponding in the same way to the other possible courses can, then, be regarded as factors or coefficients by which the distance run is to be multiplied to give the northing (or southing) and the easting (or westing). They will obviously all be fractions ranging between 0 and 1.

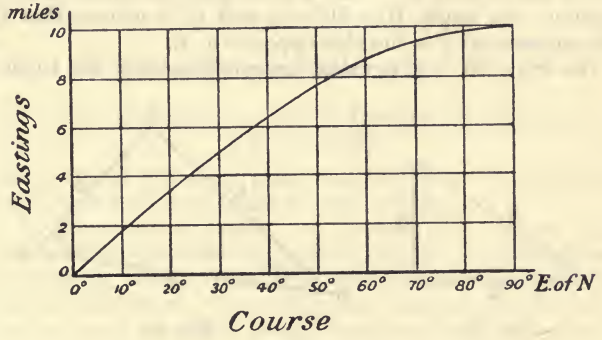


Fig. 29.

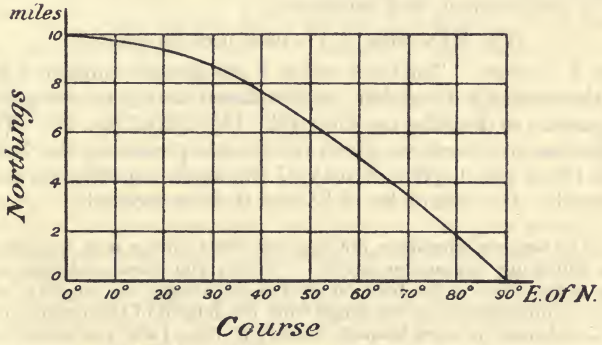


Fig. 30.

It is convenient to give these coefficients definite names. The coefficient used to calculate the easting or westing is called the *sine* of the angle to which it belongs, the other coefficient the *cosine*. They may, of course, be used to calculate the sides of any right-angled triangle when the hypoten-

use and one of the angles are given. The sine is the factor by which the hypotenuse must be multiplied to give the side *opposite* to this angle; the cosine is the factor for calculating the side *adjacent* to this angle (fig. 31).[1]

§ 4. *Complementary Angles.*—It will be noticed that the sine and cosine of an angle α are respectively the same as the cosine and sine of $90° - \alpha$. Fig. 32 shows the reason of this relation: the angle $R = 90° - \alpha$ and it is evident that the side adjacent to P is the side opposite to R.

The angle $90° - \alpha$ is called the *complement* of the angle α.

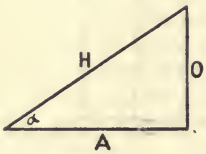


Fig. 31.

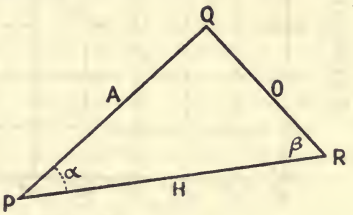


Fig. 32.

We have, then, that the sine of an angle is equal to the cosine of its complement, and conversely.

[Ex. XIX, Nos. 1-17, may now be taken.]

§ 5. *Vectors.*—The term *vector*, if not already familiar, may, at the teacher's discretion, be introduced during the foregoing argument to describe the lines OP, PQ, QR of fig. 28. The definition of a vector is given in the note preceding Ex. XIX, No. 18, at which point it may, if the teacher prefers, be considered. The rest of Ex. XIX can then be worked.

[1] The teacher who does not despise these things may be glad of the following mnemonic device: "When you were told that the *sine* is the name of the fraction used in calculating the length of the side (O) opposite the given angle from the length of the hypotenuse (H), and that the word is spelt, not s i g n, but s i n e, you might well have expressed your surprise by saying OH! When you were told that the length of the adjacent side (A) is calculated from that of the hypotenuse (H) by means of another fraction called the *co*-sine you might have made the milder remark AH! Now 'OH' comes necessarily before 'AH' and 'sine' necessarily before '*co*-sine'; thus you can always remember which stands for which."

CHAPTER XIV.

THE TRIGONOMETRICAL RATIOS (II).

A. Circles of Latitude. Middle Latitude Sailing (ch. XI., § 2; Ex. XX).

[The teacher will need a globe—preferably one with a surface upon which lines may be drawn with chalk.]

§ 1. *Changes of Latitude: the Nautical Mile.*—In ch. XIII., B, the northing (or southing) and the easting (or westing) were calculated in miles. The sailor does not actually leave his results in this form, for, as every one knows, positions on a map or chart are fixed not by measurements representing *miles*, but by measurements representing *latitude* and *longitude*. Thus, before he can use the results of his calculation the sailor must express his northing or southing in degrees and minutes of latitude and his easting or westing in degrees and minutes of longitude.

There is no difficulty with regard to the northing or southing. Reference to the globe shows that the *meridians* or circles of longitude are all "great circles" whose circumferences are simply the circumference of the earth. When we know the length of this circumference a simple calculation will give the change of latitude produced by a given northing or southing. The polar circumference of the earth is 24,856 miles, whence the length of a degree of latitude is 24856/360 = 69 miles.

Now if this number were 60 the sailor's calculations would be much simplified; every mile run north or south along a meridian would mean a difference of latitude of exactly one minute. To secure the advantage of so simple a relation the sailor abandons the ordinary or statute mile and uses a 'nautical" or sea-mile which contains 6080 feet instead of 5280. Sixty of these miles are of the same length as 69 statute miles. In other words, 60 sea-miles are equivalent to a degree of latitude and 1 sea-mile to a minute of latitude.

Thus a northing of 72 sea-miles implies a difference of latitude of 1° 12′; a northing of 354 sea-miles a difference of latitude of 5° 54′.

§ 2. *Changes of Longitude.*—The question of the difference of longitude produced by a given easting or westing is not so simple. The parallels of latitude are "small circles" whose circumferences decrease from the equator towards the poles. Only along the equator itself will 1 sea-mile imply 1′ difference of longitude. The readiest way to see this is to note how the meridians on the globe approach one another towards the poles.

What is the law which the length of a degree of longitude follows? It will probably be suggested that the length is proportional to the distance from the pole. This suggestion is easily found to be inadequate; measurement on the globe shows that the greatest distance between two meridians is halved not at latitude 45° but at latitude 60°. The explanation of this fact must be sought in the law which fixes the relative lengths of the circles of latitude. Why has the parallel of 60° half the circumference of the equator?

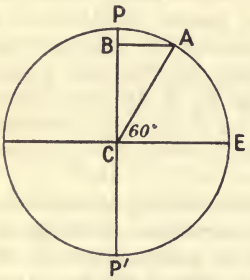


FIG. 33.

Fig. 33 answers the question. It represents a section through the centre of the globe (C) and the poles (P, P′). CE is a radius joining the centre to a point E on the equator; A is a point in latitude 60°. Let the figure rotate about PP′. The circle would trace out the surface of the globe, E the equator, A the parallel of 60°. It is evident that CE would be the radius of the equator and AB the radius of the 60th parallel. Let CE = CA = R, then

$$\begin{aligned} AB &= AC \sin ACB \\ &= R \cos 60° \\ &= 0{\cdot}5\ R, \quad \text{by the table p. 111.} \end{aligned}$$

But if the radius of the 60th parallel is one-half of the radius of the equator, its circumference will be one-half of the length of the equator. That is, the length of a degree in 60° latitude is, as we found by measurement, one-half of the length of a degree along the equator. Conversely a voyage of a given number of sea-miles along the 60° parallel implies a change of

longitude twice as great as if it had been taken along the equator.

The rule is easily generalized. If the latitude be called λ the radius of the parallel is $R \cos \lambda$. To find the change of longitude corresponding to a given easting or westing we have the rule :—

$$\text{change of longitude in minutes} = \frac{\text{no. of sea-miles}}{\cos \lambda}$$

Thus if a ship sails 420 miles along the 53rd parallel the change of longitude is $420/\cos 53^\circ = 700$ minutes $= 11^\circ 40'$.

§ 3. *Middle Latitude Sailing.*—A new difficulty now comes into view. Suppose a ship to start in longitude 42° 18′ W. and latitude 60° N. and to run 20 miles on a course 37° E. of N., then we should have :—

$$\text{northing} = 20 \cos 37^\circ = 16 \text{ miles} \qquad \text{easting} = 20 \sin 37^\circ = 12 \text{ miles}$$

The 16 sea-miles to the north imply a difference of latitude of 16 minutes. The ship's latitude is, therefore, now 60° 16′ N. But when we seek to convert the easting into minutes of longitude what cosine are we to take as divisor? If the ship had sailed 12 miles east in latitude 60° the divisor would have been cos 60° ; if in latitude 60° 16′ it would have been cos 60° 16′. But as a matter of fact it made its easting on neither of these parallels but on its way from one to the other. A strictly accurate result cannot be obtained, therefore, by using either of the cosines as divisor.

This objection is sound, but in the case before us is not serious. There is so little difference between the two cosines that it hardly matters which we take as divisor. We can without serious error assume that all the easting took place in latitude 60°. Then we have :—

$$\text{diff. of long.} = 12/\cos 60^\circ = 24'$$

so that the new longitude is $42^\circ 18' - 24' = 41^\circ 54'$ W.

But now suppose the distance run to have been 200 miles, so that the northing would be 160 miles and the easting 120 miles. In the first place we have :—

$$\text{change of lat.} = 160' = 2^\circ 40'$$
$$\therefore \text{new latitude} = 62^\circ 40'$$

The cosine graph shows that cos 62° 40′ = 0·46. The difference between

120/cos 62° 40′ and 120/cos 60° is 261′ − 240′ = 21′.

This is too serious a difference to be ignored; so the sailor compromises and assumes that the easting was made on a parallel *half-way between* 60° and 62° 40′, i.e. 61° 20′. The graph shows that cos 61° 20′ = 0·48. So we have:—

$$\begin{aligned} \text{change of long.} &= 120/0{\cdot}48 \\ &= 250' \\ &= 4^\circ\ 10' \\ \therefore \text{ new longitude} &= 42^\circ\ 18' - 4^\circ\ 10' \\ &= 38^\circ\ 8'\ \text{W.} \end{aligned}$$

This method is called the rule of *middle latitude sailing*.

The question may be asked—Would the rule apply if the distance run were much greater still—say 2000 miles? The answer is obviously, No. It can only be used for moderate runs. But as a single day's run is never anything like so great as 2000 miles the failure of the rule is unimportant. The final longitude after several days' running can always be determined by adding the daily changes together.[1]

[Ex. XX may now be taken.]

B. The Relations between the Sine, Cosine and Tangent of an Angle. The Calculation of the Ratios of Certain Angles (ch. XI., § 2; Ex. XXI).

§ 1. *Tan a = sin a/cos a.*—The tables on pages 107 and 111 give the values of the tangents, sines and cosines as they could be determined by careful drawing and measurement. But when numbers are obtained by a graphic process it is always well to check them by calculation if it is possible to do so. There are several ways in which the accuracy of our tables can be tested. Here is one way. In fig. 31 we have

$$\tan a = \text{O/A},\ \sin a = \text{O/H},\ \cos a = \text{A/H}$$

$$\text{But}\quad \frac{\text{O}}{\text{A}} = \frac{\text{O}}{\text{H}} \div \frac{\text{A}}{\text{H}};$$

that is, $\tan a = \sin a/\cos a$.

If, therefore, the numbers in the tables are correct the value

[1] The complete theory of Mercator sailing will be found in Part II of this work.

given in the first as the tangent of an angle should be equal to the quotient of the sine of the same angle by the cosine, as these are given in the second table.

§ 2. *Sin*$^2\,\alpha + cos^2\,\alpha = 1$.—A second test applies only to the sine and cosine of the same angle. Let the triangle ABC (fig. 34) be right angled at C, and let the perpendicular CD divide the base into two parts of length p and q. Then we have : $b = c\cos\alpha$, $a = c\cos\beta = c\sin\alpha$, and :—

$$\begin{aligned} c &= p + q \\ &= b\cos\alpha + a\cos\beta \\ &= b\cos\alpha + a\sin\alpha \\ &= c\cos^2\alpha + c\sin^2\alpha \\ &= c\,(\cos^2\alpha + \sin^2\alpha) \end{aligned}$$

Hence (i) $\sin^2\alpha + \cos^2\alpha = 1$, (ii) $\cos\alpha = \sqrt{1 - \sin^2\alpha}$, and (iii) $\sin\alpha = \sqrt{1 - \cos^2\alpha}$.

These formulæ are proved still more easily if the class knows

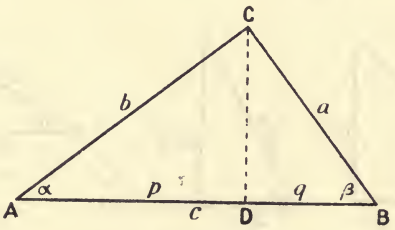


Fig. 34.

the theorem of Pythagoras[1] (Euclid I, 47), for we then have :—

$$\begin{aligned} a^2 + b^2 &= c^2 \\ a^2/c^2 + b^2/c^2 &= 1 \end{aligned}$$

that is, $\sin^2\alpha + \cos^2\alpha = 1$.

§ 3. *The Identities Combined.*—The identities of §§ 1, 2 can be combined to yield tests by which, given the sine or cosine, the tangent can be calculated, and *vice versa*. Thus we have :—

[1] This theorem can itself be proved most easily by fig. 34. We have :—

$$\cos\alpha = b/c = p/b\ ;\ \sin\alpha = \cos\beta = a/c = q/a$$
$$\therefore\ b^2 = pc \qquad\qquad \therefore\ a^2 = qc$$
$$\text{whence } a^2 + b^2 = (p + q)c = c^2.$$

$$\tan \alpha = \sin \alpha / \cos \alpha$$
$$= \sin \alpha / \sqrt{(1 - \sin^2 \alpha)} \text{ by § 2, (ii)}$$
$$= \sqrt{(1 - \cos^2 \alpha)} / \cos \alpha \text{ by § 2, (iii)}$$

Again
$$\tan^2 \alpha + 1 = \frac{\sin^2 \alpha}{\cos^2 \alpha} + 1$$
$$= \frac{\sin^2 \alpha + \cos^2 \alpha}{\cos^2 \alpha}$$
$$= 1/\cos^2 \alpha$$

whence
$$\cos \alpha = 1/\sqrt{(\tan^2 \alpha + 1)}$$

Finally
$$\sin \alpha = \cos \alpha \frac{\sin \alpha}{\cos \alpha}$$
$$= \cos \alpha \tan \alpha$$
$$= \tan \alpha / \sqrt{(\tan^2 \alpha + 1)}$$

by the preceding result.

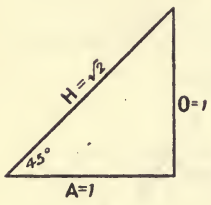


Fig. 35.

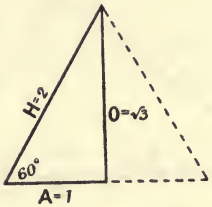


Fig. 36.

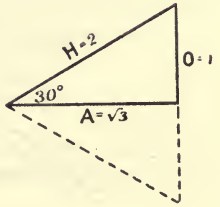


Fig. 37.

§ 4. *The Ratios of* 45°, 30°, 60°.—The preceding tests apply to all angles. In the case of certain special angles still better tests are available. Thus if $\alpha = 45°$ and we put $A = 1$ (fig. 35), it follows (since the triangle is isosceles) that $O = 1$. By Pythagoras' Theorem (§ 2 above)

$$H = \sqrt{A^2 + O^2} = \sqrt{2} = 1{\cdot}414.$$

Hence $\tan 45° = 1$, $\sin 45° = \cos 45° = 1/\sqrt{2} = 0{\cdot}707$.

All three results are in agreement with the tables.

In fig. 36 the whole triangle is an equilateral triangle which has been divided into two identical right-angled triangles by a perpendicular from the vertex to the base. Here, if we put $H = 2$, we have $A = 1$ and $O = \sqrt{H^2 - A^2} = \sqrt{3}$.

Hence tan $60° = \sqrt{3} = 1{\cdot}732$, sin $60° = \sqrt{3}/2 = 0{\cdot}866$, cos $60° = 1/2 = 0{\cdot}500$. Fig. 37 shows the equilateral triangle differently arranged, and gives us tan $30° = 1/\sqrt{3} = 0{\cdot}577$, sin $30° = 0{\cdot}500$, cos $30° = 0{\cdot}866$. All these six results agree with those in the tables.

[Ex. XXI should now be taken.]

CHAPTER XV.

THE COMBINING OF FORMULÆ.

A. The Determination of Constants in a Formula (ch. XI., § 3; Ex. XXII, A).

§ 1. *Determination by a Graph.*—When a weight of 20 grm. is suspended by a rubber cord the length of the cord is 25·1 cm. When the weight is 30 grm. the length is 28·5 cm. Assuming a linear relation between l and w find the formula.

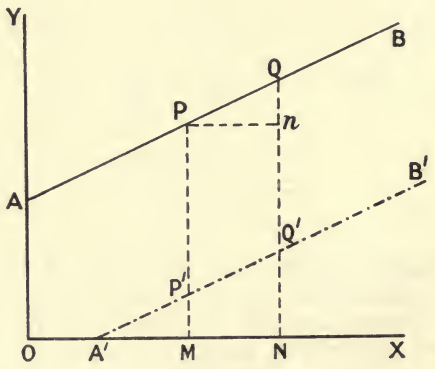


FIG. 38.

This problem is easily solved by the methods of Ex. XVII, Nos. 18, 19, 20. Let AB (fig. 38) be the graph, P and Q the points representing the data. Then $Qn = 28{\cdot}5 - 25{\cdot}1 = 3{\cdot}4$ cm. is the increase of length for the addition of

$$Pn = 30 - 20 = 10 \text{ grm.}$$

Thus the increase of length per gramme added is

$$Qn/Pn = 0{\cdot}34$$

To find the original length AO we must subtract from PM

$$0{\cdot}34 \times 20 = 6{\cdot}8 \text{ cm.}$$

Hence

$$\text{AO} = 25{\cdot}1 - 6{\cdot}8 = 18{\cdot}3 \text{ cm.}$$

Thus the formula is

$$l = 18{\cdot}3 + 0{\cdot}34w$$

The method can, of course, be generalized. Let the linear relation between any variables x and y have the form

$$y = a + bx$$

where a and b are constant numbers. To fix the exact form of the relation in a given case a and b must be known. They can be determined if two pairs of corresponding values of x and y are given. For example, let the pairs be $x = 8$, $y = 41$, and $x = 13$, $y = 56$. Then $\text{Q}n = 56 - 41 = 15$, $\text{P}n = 13 - 8 = 5$ and $\text{Q}n/\text{P}n = 15/5 = 3$. Also

$$\text{AO} = \text{PM} - 3 \times 8 = 41 - 24 = 17.$$

Hence

$$y = 17 + 3x.$$

§ 2. *Algebraic Methods:* (i) *by Composition.*—Let us now seek rules for determining a and b without drawing a figure. We have

$$y = a + bx$$

$$(\text{QN})\ 56 = a + 13b \quad . \quad . \quad . \quad \text{(i)}$$

$$(\text{PM})\ 41 = a + 8b \quad . \quad . \quad . \quad \text{(ii)}$$

$$(\text{Q}n)\ 15 = 5b^{1} \quad . \quad . \quad . \quad . \quad \text{(iii)}$$

whence $b = 3$

Again by (ii)

$$41 = a + 8b$$
$$= a + 24 \text{ (since } b = 3)$$

$$(\text{AO})\ a = 41 - 24$$
$$= 17$$

Thus the relation is $y = 17 + 3x$.

The essence of the method is that by subtracting the two sides of (ii) from the corresponding sides of (i) we obtain a new relation (iii) from which the value of one constant is at once determined. This known, the value of the other can be obtained from one of the original relations. We may call this *the method of composition,* since relation (iii) is obtained by compounding (i) and (ii).

§ 3. *Algebraic Methods:* (ii) *by Substitution.*—Another method is readily suggested. In (i) change the subject to a:—

$$a = 56 - 13b \quad . \quad . \quad . \quad . \quad \text{(iv)}$$

[1] Note that $5 = \text{P}n$.

In (ii) replace a by its equivalent in (iv). Thus

$$41 = (56 - 13b) + 8b$$

whence $b = 3$. The value of a may be determined as before.

This method is called the *substitution* method.

§ 4. *The Relation* $y = a - bx$.—A linear relation is not always of the form $y = a + bx$. If y decreases as x increases a graph shows that the form must be $y = a - bx$. For example, let (5, 33) and (12, 5) be the two pairs of values of x and y. Then, by the composition method :—

$$\begin{aligned} y &= a - bx \\ 33 &= a - 5b \quad . \quad . \quad . \quad . \quad . \quad \text{(i)} \\ 5 &= a - 12b \quad . \quad . \quad . \quad . \quad \text{(ii)} \\ 33 - 5 &= (a - 5b) - (a - 12b) \\ \text{or} \qquad 28 &= 7b \quad . \quad . \quad . \quad . \quad . \quad \text{(iii)} \\ b &= 4 \end{aligned}$$

The line between (ii) and (iii) may be omitted when its effect is understood.

The substitution method gives the same result for b ; a can be determined as in § 2.

§ 5. *The Relation* $y = bx - a$.—Consider now the relation $y = 4x - 12$. Its graph must be a straight line, for if x increases by equal steps y increases also by equal steps. But it is clear that $4x$ can never be less than 12, i.e. x cannot be less than 3. When $x = 3$, $y = 0$. Thus $y = 4x - 12$ is a linear relation whose graph begins, like A′B′ (fig. 38), at some point on the axis of x instead of beginning, like AB, at some point on the axis of y. Relations corresponding to lines in such positions must be of the form $y = bx - a$.

Is there any way of finding, without drawing the graph, which form the linear relation has in a given case? Consider the relations :—

$$y = bx, \qquad y = bx + a, \qquad y = bx - a.$$

These may be written :—

$$y/x = b, \qquad y/x = b + a/x, \qquad y/x = b - a/x.$$

In the first the quotient y/x is the same for all values of x ; in the second it decreases as x increases, in the third it increases as x increases.[1]

[1] For the fraction a/x grows smaller as its denominator x grows larger. Thus as x increases there is a smaller amount to add in the second case and to subtract in the third case.

We can, then, decide the relation appropriate to a given pair of values of x and y by noting the value of y/x for each pair. If this value decreases with a greater value of x the relation is $y = a + bx$; if it increases the relation is $y = bx - a$. For example, find by the composition method a linear relation such that when $x = 7$, $y = 33$, and when $x = 11$, $y = 57$. Here $57/11 > 33/7$ so that the relation is

$$\begin{aligned} y &= bx - a \\ 57 &= 11b - a \quad . \quad . \quad . \quad . \quad . \quad \text{(i)} \\ 33 &= 7b - a \quad . \quad . \quad . \quad . \quad \text{(ii)} \\ 57 - 33 &= (11b - a) - (7b - a) \\ 24 &= 4b \quad . \quad . \quad . \quad . \quad . \quad \text{(iii)} \\ b &= 6 \end{aligned}$$

From (i) or (ii) $a = 9$. Hence $y = 6x - 9$.

[Ex. XXII, A, may now be worked.]

B. Common Values of Two Relations (ch. XI., § 3; Ex. XXII, B).

§ 1. *Determination of Common Values by Graph.*—Two rubber cords hang side by side and are loaded with weights as in A, § 1. The formula giving the length of one cord is

$$l = 14{\cdot}2 + 1{\cdot}2w;$$

the formula for the other cord is

$$l = 18{\cdot}7 + 0{\cdot}7w.$$

The units of measurement are inches and ounces. Is there any weight which will make the cords assume the same length?

This problem is similar to Ex. XVII, Nos. 21-24. Let AB (fig. 39) be the graph of the first formula, A′B′ that of the second. Then C clearly gives a weight OM which will stretch each cord to the same length, CM. To find this weight by calculation we have :—

$$\begin{aligned} \text{CM} &= \text{AO} + \text{C}m \quad \text{and also} \quad \text{CM} = \text{A}'\text{O} + \text{C}m' \\ &= 14{\cdot}2 + 1{\cdot}2w \qquad\qquad = 18{\cdot}7 + 0{\cdot}7w \end{aligned}$$

Hence we can put

$$\begin{aligned} 14{\cdot}2 + 1{\cdot}2w &= 18{\cdot}7 + 0{\cdot}7w \\ w &= 9 \text{ ounces.} \end{aligned}$$

From either of the original relations we now find $l = 25$ inches.

§ 2. *Algebraic Methods.*—The method just employed can obviously be regarded as the "substitution method" employed to find the value of a variable instead of the value of a constant. For example, let the question be proposed whether the two linear relations $y = 14 + 6x$ and $y = 8x - 7$ have a pair of values of x and y in common. By the method of substitution we can replace y in the second relation by $14 + 6x$. We then have

$$14 + 6x = 8x - 7$$
$$x = 10{\cdot}5$$

From the first relation it now follows that $y = 77$. Thus the pair (10·5, 77) is common to both relations.

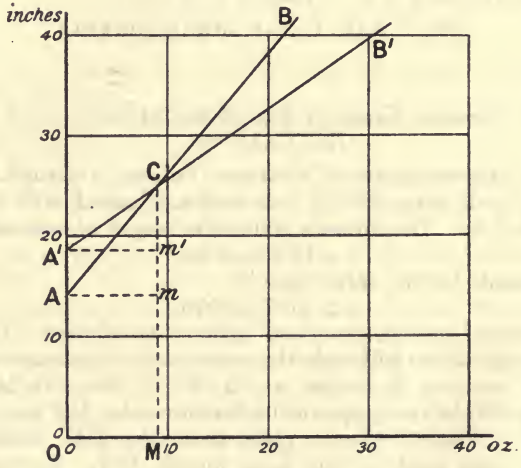


FIG. 39.

Have the relations $3x - 4y = 7$ and $5x + 2y = 16$ any common pairs of values of x and y?

The relations can be written $y = \frac{3}{4}x - \frac{7}{4}$ and $y = 8 - \frac{5}{2}x$ respectively. These transformations show that they are linear so that they may have one pair of values in common. Using the substitution method we have :—

$$\frac{3}{4}x - \frac{7}{4} = 8 - \frac{5}{2}x$$

$$\frac{13}{4}x = \frac{39}{4}$$

$$x = 3$$

From $y = \frac{3}{4}x - \frac{7}{4}$ we now have $y = \frac{1}{2}$.

The transformation that necessarily precedes substitution in such examples as this could be avoided by a " composition " method. We have :—

$$3x - 4y = 7 \quad . \quad . \quad . \quad \text{(i)}$$
$$5x + 2y = 16 \quad . \quad . \quad . \quad \text{(ii)}$$

From (ii) we have

$$10x + 4y = 32 \quad . \quad . \quad . \quad \text{(iii)}$$

If (i) and (iii) be now added y will disappear and we shall have :—

$$13x = 39 \quad . \quad . \quad . \quad \text{(iv)}$$
$$x = 3$$

From (i) $\quad 9 - 4y = 7$

$$y = \frac{1}{2}$$

[The class should at this point turn to Ex. XXII, No. 13, and should (without working out the examples) consider how the composition method is to be applied to them.]

§ 3. *Common Values not always Possible.*—It is evident that if two graphs do not intersect the corresponding relations have no pair of common values of x and y. A little consideration will determine without the aid of graphs whether common values do or do not exist in a given case.

Have the relations $2x - 3y = 4$ and $3y - 5x = 7$ a common pair of values of x and y? Here, using the method of composition, we have :—

$$2x - 3y = 4 \quad . \quad . \quad . \quad . \quad \text{(i)}$$
$$3y - 5x = 7 \quad . \quad . \quad . \quad . \quad \text{(ii)}$$
$$2x - 5x = 4 + 7 \quad . \quad . \quad . \quad \text{(iii)}$$

But, as $5x$ cannot be taken from $2x$, (iii) cannot be formed.[1] Thus no single pair of values of x and y satisfies both relations, and the graphs will be found not to cross.

[1] Note that $2x - 5x = 4 - 7$ would give no difficulty for it could be transformed into $7 - 4 = 5x - 2x$.

Consider next the relations $3x - 4y = 5$ and $4x - 3y = 7$. Multiplying these by 4 and 3 respectively we obtain :—

$$12x - 16y = 20 \quad . \quad . \quad . \quad \text{(i)}$$
$$12x - 9y = 21 \quad . \quad . \quad . \quad \text{(ii)}$$

Since $21 > 20$ we subtract downwards and obtain

$$16y - 9y = 1$$
$$y = \tfrac{1}{7}$$

and from $3x - 4y = 5$ obtain $x = 1\tfrac{6}{7}$.

But if the former relation had been $3x - 4y = 6$ we should have had

$$12x - 16y = 24$$
$$12x - 9y = 21$$

Here subtraction of the lower line from the upper leads to

$$9y - 16y = 24 - 21$$

which is obviously impossible. There is, then, no common pair of values of x and y.

Finally take the pair $y = \frac{3}{2}x + 6$ and $y = \frac{7}{4} - \frac{5}{8}x$. The substitution method gives :—

$$\tfrac{3}{2}x + 6 = \tfrac{7}{4} - \tfrac{5}{8}x$$

leading to $\quad \tfrac{3}{2}x + \tfrac{5}{8}x = \tfrac{7}{4} - 6$

which is impossible.

On the other hand, suppose the relations to have been given in the equivalent forms $2y - 3x = 12$ and $5x + 8y = 14$. Applying the composition method we should have, after multiplying the former by 5 and the latter by 3 :—

$$10y - 15x = 60 \quad . \quad . \quad . \quad \text{(i)}$$
$$15x + 24y = 42 \quad . \quad . \quad . \quad \text{(ii)}$$

and, by addition,

$$34y = 102$$
$$y = 3$$

But on putting $y = 3$ in $5x + 8y = 14$ we have

$$5x + 24 = 14$$

which is again impossible. Thus we cannot, by either method, find a pair of values of x and y common to the two relations.

[Ex. XXII, B, may now be worked.]

C. Elimination[1] (ch. XI., § 3 ; Ex. XXII, C).

§ 1. *An Example of Elimination.*—Turn back to Ex. XVIII, No. 13 ; let h be the height of the flagstaff and d the distance

[1] This section may be omitted and taken in revision.

of the observer from the building. Then we have

$$(h + 110)/d = \tan 46^\circ \quad . \quad . \quad . \quad \text{(i)}$$

Now this relation, as it stands, does not enable us to calculate h because of the presence of d. We must find, therefore, some means of *eliminating* or getting rid of d. This can be done by the help of the relation $110/d = \tan 37^\circ$, using either the method of substitution or the method of composition. By the former method we have $d = 110/\tan 37^\circ$, which, substituted in (i), leads to

$$(h + 110) \tan 37^\circ/110 = \tan 46^\circ \quad . \quad . \quad \text{(ii)}$$

Composition must in this case take the form not of adding or subtracting but of dividing the first relation by the second. Thus we have :—

$$\frac{h + 110}{d} \times \frac{d}{110} = \tan 46^\circ \times \frac{1}{\tan 37^\circ}$$

or

$$(h + 110)/110 = \tan 46^\circ/\tan 37^\circ \quad . \quad . \quad \text{(iii)}$$

—a relation obviously equivalent to (ii). From either of them it is easy to calculate h.

§ 2. *Elimination in General.*—The foregoing problem affords an example of the process called *elimination*. Elimination, in general, aims at reducing the number of variables or unknown numbers in a given set of relations. Thus in § 1 our aim was to obtain a relation involving h only from two relations involving h and d. The following is an example in which the number of variables, originally three, is reduced to two.

Eliminate z from the relations

$$3x - 4z = 13 \text{ and } 5y + 2z = 3.$$

Upon multiplying the second relation by 2 and compounding, z disappears, and we have

$$3x + 10y = 19$$

as the required relation containing only x and y.

§ 3. *Trigonometrical Examples.*—The term composition is to be taken to mean any manipulation of two given relations which leads to a third relation different from either of them. The variety of manipulations sometimes required in elimination may be illustrated by two examples.

(i) Eliminate α from the three relations

$$x/a = \sin \alpha, \; b/y = \cos \alpha, \; c/xy = \tan \alpha.$$

Here we have, by compounding the first and second rela-

tions, $xy/ab = \sin\alpha/\cos\alpha = \tan\alpha$. Substitution for $\tan\alpha$ in the third relation leads to

$$xy/ab = c/xy \qquad \text{or} \qquad x^2y^2 = abc.$$

(ii) Eliminate α from the relations

$$x = a\cos\alpha,\ y = b\sin\alpha$$

Here we argue as follows :—

$$x = a\cos\alpha \qquad \therefore \cos\alpha = x/a$$
$$y = b\sin\alpha \qquad \therefore \sin\alpha = y/b$$

But $\cos^2\alpha + \sin^2\alpha = 1$; $\quad \therefore x^2/a^2 + y^2/b^2 = 1.$

[Ex. XXII, C, may now be worked.]

CHAPTER XVI.

FURTHER TYPES OF PROPORTIONALITY.

A. Inverse Proportion (ch. IV., § 8; ch. XI., § 1; Ex. XXIII).

§ 1. *Inverse Proportion; its Graphic Symbol.*—Two places, P and Q, are 16 miles apart. How long will it take to travel with constant speed from P to Q? The answer obviously depends upon the rate of movement. A walker going at 4 miles/hour will take four hours; a donkey-cart going twice as fast will take half as long; a cyclist going three times as fast one-third as long; an express train going 40 miles/hour, one-tenth as long, and so on. These facts are expressed by saying that the time taken is *inversely proportional* to the speed.

We saw that direct proportion has a definite graphic symbol—the straight line through the origin. Let us see what results from graphing a series of pairs of numbers in inverse proportion. The problem just considered will serve as an example. The speed of the movement must be taken as the independent variable. Any number of points may readily be determined by noting that the product of speed and time must always be 16. Some small speeds should be considered (e.g. a creature crawling $\frac{1}{2}$ ml./hr.) as well as large speeds. The points obtained determine the smooth curve A (fig. 40).

Now if A can really be taken as a graphic symbol of this case of inverse proportion then the product of the co-ordinates of *any* point on the curve should be 16, just as the ratio of any pair of co-ordinates in the straight line symbol of direct proportion is k. In the case of the straight line we were able to prove the required property by geometry. Since we do not know any properties of the curve A from which to start a proof by reasoning we must be contented with the results of measure-

ment. Let each pupil choose, at random, a couple of points on the curve and find that the product of the co-ordinates is, in every case, as nearly 16 as we could expect. We may then conclude that the curve A is truly the graphic symbol of the formula $st = 16$, and, when once drawn, may be used (as the straight lines were used in ch. XII., A) as a ready-reckoner. [Examples of interpolation should be given.]

§ 2. *Generalization of Formula and Graphic Symbol.*—As in ch. XII., we may now note that the curve A would be the graphic symbol of *any* case of inverse proportion between two variables whose constant product is 16. That is, we may regard the curve as the graph corresponding to the relation

$$xy = 16 \quad \text{or} \quad y = 16/x$$

where x and y represent any variables whatever.

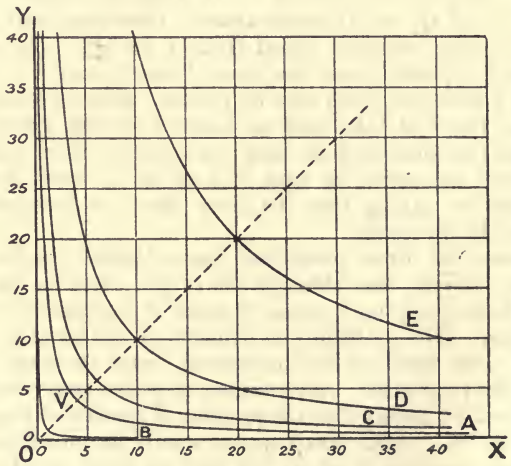


FIG. 40.

It is evident that the 16 here plays a part similar to the constant, k, in the formula $y = kx$. Now when k received different values we found that the corresponding graphs were all straight lines, and that they differed only in their inclination to the axis of x. It is important, therefore, to inquire how the graphic symbol of inverse proportion will change if we substitute another constant for 16. In other

words, what are the graphs of the relation $y = k/x$ corresponding to different values of k?

The investigation is best divided among the class, each taking a different value for k. Some of the curves should be transferred with rough accuracy to the blackboard. It will then be seen that we have a number of curves which all bear a family resemblance to A. For example, B (fig. 40) is the curve answering to $k = 1$, C, D and E those obtained when $k = 36$, $k = 100$ and $k = 400$. The "family resemblance" consists in the following facts. (i) The straight line, OV, drawn at an inclination of 45° to OX, divides each curve symmetrically; for every point on one side of this line there is a corresponding point on the other. The line OV is for this reason called the *axis* of the curves. (ii) The point where this line cuts each curve is the point nearest to the origin. It is called the *vertex* or head of the curve. (iii) At any vertex, V, x and y are equal; and, since $xy = k$, we have $x = y = \sqrt{k}$. Hence the distance

$$\mathrm{OV} = \sqrt{(x^2 + y^2)} = \sqrt{2k}$$

(iv) Each curve approaches constantly nearer to the axes but never reaches them. However large x is in $y = k/x$, y can never be zero; and, however small x is, it will always be possible to calculate a value of y though it will become endlessly large. Thus neither x nor y can be ever absolutely zero. These facts are expressed by saying that the axes are *asymptotes* of the curves.

It is evident, then, that we can regard all the curves of fig. 40 as simply different specimens of the same kind of curve, just as the curves we draw with compasses are all specimens of the circle. We will call it the *inverse proportion curve.*[1]

§ 3. *Mechanical Constructions for the Curve.*[2]—When we need the graph corresponding to $y = kx$ it is unnecessary to plot a number of points. Knowing that it will be a straight line through the origin we determine a single point and join this point up with the origin by means of the ruler. It may be asked whether there is any means of drawing the curve corresponding to $y = k/x$ without plotting points, just as a line is drawn with a ruler or a circle with compasses.

[1] The term "rectangular hyperbola" is, perhaps, best reserved until ch. XXVI. is reached.

[2] This section may be omitted or taken in revision.

Such means exist though they are not so simple as the use of ruler or compasses. The following is, perhaps, the simplest.

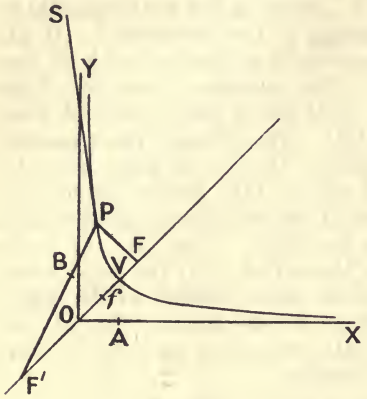


FIG. 41.

Begin by drawing the axis F'OF (fig. 41). On OX take A where $OA = \sqrt{k}$, and so determine the vertex V. Mark off OF and OF' each of length $2\sqrt{k}$. Also determine a point f where $Vf = VF$. Now take a piece of string, tape or other unstretchable material (F'B) and fasten to one end of it a length of elastic cord. Lay the combination along F'F so that (B) the join of string and elastic cord is at f. Fix the string by a drawing-pin at F' and the elastic cord by another at F. Fasten a piece of string, SP, to the elastic cord at V and pull the string out gradually, taking care that it always makes equal angles with the two parts of the cord. The point of a pencil following the movement of the point on the elastic cord corresponding to the point P will trace out the inverse proportion curve.[1]

A more accurate but less simple method[2] is to hold a ruler so that it rotates about F'. At the other end fix on its edge at K a piece of string which is shorter than KF' by a length F'f. Fasten the free end at F. As the ruler rotates hold the string tight against its edge by the pencil point. The curve will in this way be obtained.

[Ex. XXIII may now be worked.]

[1] In the rectangular hyperbola the eccentricity $= \sqrt{2}$. Thus if F is a focus, $OF = \sqrt{2}$, $OV = 2\sqrt{k}$. Also, if P is any point on the curve, and F' the other focus, $PF' - PF = 2OV = F'f$. Since PF = PB and since F'B does not stretch the last condition is fulfilled.

[2] Described by Milne and Davis, *Geometrical Conics*, p. 82.

B. Direct Proportion to the Square or the Square Root (ch. IV., § 8; ch. XI., § 1; Ex. XXIV).

§ 1. *Graphic Symbol of* $y = kx^2$.—Suppose a number of squares to be cut from a uniform sheet of cardboard or metal and to be weighed. The weight (W) would be directly proportional to the area (A). The formula expressing the relation is $W = kA$ and its graphic symbol a straight line through the origin. In fig. 42, the line OP corresponds to the formula in the case where $k = \frac{1}{2}$.

Now, if l is the length of the side of a square, $A = l^2$ and $W = kl^2$. The weight of a square is, then, directly proportional not only to its area but also to the *square* of its side. If we regard OP (fig. 42) as the graph of the formula $W = \frac{1}{2}l^2$ then the numbers along the horizontal axis are values not of l but of the square of l. We may ask what form the graph would take if we regarded these numbers as values of l itself.

Consider the point P, which indicates the weight of a square for which $A = 16$. For this square $l = 4$; consequently if

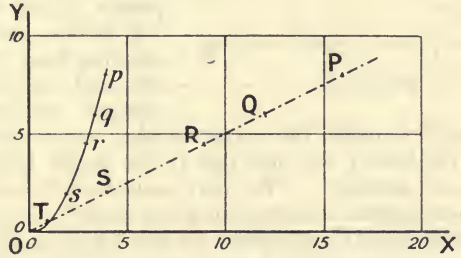


FIG. 42.

the base numbers are to be regarded as values of l instead of values of A the point P must be moved until it stands in the position p above 4. Similarly Q must be moved to q above 3·46 ($= \sqrt{12}$), R above 3 ($= \sqrt{9}$), etc. The point T above 1 will remain undisturbed, but the point above $\frac{1}{4}$ must be moved to the right till it stands above $\frac{1}{2}$.

Joining the points by a smooth curve we have Op suggested as the graphic symbol of $W = \frac{1}{2}l^2$. As in the last lesson the validity of the suggestion must be tested by taking points at

random and seeing whether their co-ordinates satisfy the relation.

Generalizing as before we may now regard Op as the graph corresponding to the formula $y = \frac{1}{2}x^2$ in which reference is made to no special pair of variables. Further, we may regard this formula as a special case of the formula $y = kx^2$ and inquire how the form of the graph changes with the values of k.

The work is, as before, best divided among the class, the results being represented with rough accuracy upon the blackboard. The points should be obtained by direct substitution in the formula. In fig. 43 the curves A, B, C, D represent the results of putting $k = \frac{1}{12}$, $\frac{1}{4}$, 1, 3. It is obvious that the curves form a family with the properties (i) of touching OX at O, and (ii) of receding from OX without end as x is increased without end.

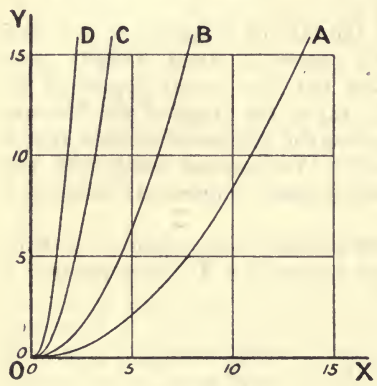


FIG. 43.

The point O is called the *vertex* of each curve. Seeking a name for the family, we note that C, the graph of $y = x^2$ is a "curve of squares". We may conveniently extend this name [1] to all graphs corresponding to $y = kx^2$.

§ 2. *Mechanical Construction for the Curve.*[2]—The curve of squares corresponding to any value of k can be drawn mechanically as follows. On OY (fig. 44) take F so that OF = $1/4k$. Draw DD′ parallel to OX and $1/4k$ below it. Take a thick set-square or any right-angled board, A. Pin at B a piece of string of length BC. Pin the free end at F. Draw the string tight by a pencil at P and slide A along DD′. The curve corresponding to $y = kx^2$ will be traced.[3]

[1] Reserving the name parabola for the complete curve studied in ch. XXVI., B.

[2] See note on A, § 3. The method described is taken from Milne and Davis, op. cit.

[3] Putting OF $\equiv a$, we have $x^2 = 4ay$; whence $a = 1/4k$.

§ 3. *The Graph of* $y = k\sqrt{x}$.—The distance visible from the top of a cliff is given (Ex. IV, No. 6) by the formula $d = 1{\cdot}22\sqrt{h}$. This is a case of direct proportion to the square root of a variable. We may obtain the corresponding graph as a straight line through the origin if we take the *square root* of h as the independent variable. But if we now wish to exhibit h itself as the independent variable each point in the straight line must be moved horizontally until it stands over the *square* of the original abscissa. For if the abscissa of a point P was 4 when the abscissa represented $\sqrt{h}$ it must be taken as 16 when it represents h itself.

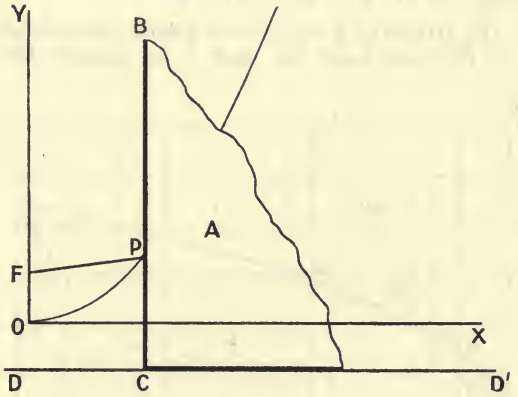


Fig. 44.

The transformation from one set of abscissæ to the other is shown in fig. 45. The points P, Q of the straight line which stand over the numbers 3 and 2 are removed horizontally to p, q where they stand over 9 and 4. The point R standing over 1 is undisturbed. The point S standing over $\frac{1}{2}$ is moved to the *left* and stands at s over $\frac{1}{4}$.

The curve Op can now be regarded as merely a special example of the graph of $y = k\sqrt{x}$ in which $k = 1{\cdot}22$. If desired we can, as before, explore the results that follow from assigning different values to k. The different curves can be drawn in three ways. (i) The co-ordinates of the points can be obtained by direct substitution in the formula. (ii) Draw the line $y = kx$ and shift the points horizontally until they

stand above the squares of the original numbers. (iii) Take the graph of $y = x^2/k^2$, turn the paper in its own plane through a right angle in the clockwise direction; turn it upside down, transferring the graph to the back of the paper by tracing against a window or by pricking points.

The graph of $y = k\sqrt{x}$ may be called the *square root curve.*

[Ex. XXIV, A, may now be worked.]

C. Inverse Proportion to the Square or the Square Root of a Variable (ch. XI., § 1; Ex. XXIV, B).

§ 1. *The Graph of* $y = k/x^2$.—A number of tins are to be made of different sizes but each is to contain the same

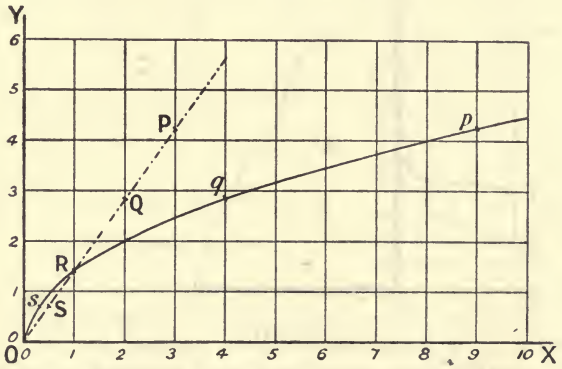


FIG. 45.

quantity of liquid. The height h of any tin is given by $h = k/A$, where k is the constant volume and A the area of the bottom. That is, the height is inversely proportional to the area of the bottom. The graph, when $k = 4\frac{1}{2}$ cu. feet and h and A are measured in linear and square feet, is the curve PTU (fig. 46).

Suppose the bottom of each tin to be a square. Then the formula becomes $h = k/l^2$, that is, the height is inversely proportional to the square of the side of the base. In order that the graph may express the relation in this form, the abscissæ must be taken to represent l instead of A and the points P, Q, R,

etc., must be removed horizontally until they stand at p, q, r, etc., above the square roots of the original numbers. (Cf. B, § 1.)

Generalizing, we may say that $p\mathrm{T}u$ is the graph of the relation $y = k/x^2$ when $k = 4{\cdot}5$. Similar curves would be obtained by assigning other values to k.

§ 2. *The Graph of* $y = k/\sqrt{x}$.—In exactly the same way we could obtain the graph corresponding to the relation $y = k/\sqrt{x}$—*i.e.* the connexion between two variables one of

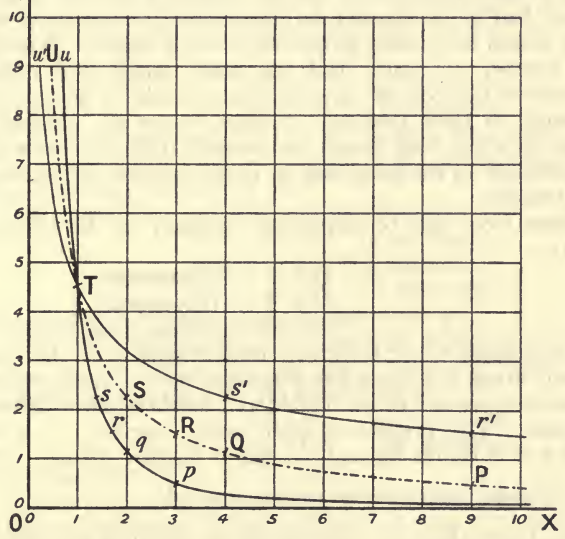


FIG. 46.

which is inversely proportional to the square root of the other. The case when $k = 4{\cdot}5$ is obtained by moving the points P, Q, R, S, etc., of fig. 46 horizontally to the positions p', q', r', s', etc., where they stand above numbers which are the squares of the original abscissæ. Ex. IV, No. 31, gives a simple instance of this kind of proportion.

These methods of obtaining the graphs of $y = k/x^2$ and $y = k/\sqrt{x}$ are instructive as indicating their relations to the inverse proportion curve. The graphs could, of course, also

be plotted as the result of direct substitution in the formulæ. Moreover the graph of $y = k/\sqrt{x}$ could be derived from that of $y = 1/k^2x^2$ in the manner described in B, § 3.

[Ex. XXIV, B, may now be worked.]

D. Combinations of Types of Proportion (ch. XI., § 1; Ex. XXV).

§ 1. *Conjoint Direct Variation.*—A suite of rooms has the same breadth throughout, but the lengths of the rooms differ. If they had to be carpeted we should say that the cost of each room would be directly proportional to the length. Similarly if a number of rooms had the same length but differed in breadth the cost of carpeting them would be directly proportional to their breadth. Finally, if we are considering rooms in which both length and breadth differ the cost will be proportional to the area, that is, to the *product* of the length and breadth.

These facts can be expressed concisely in the following form:—

$C \propto l$ (b constant)
$C \propto b$ (l constant)
$C \propto bl$ (when b and l vary)

The symbol $\propto$ $\equiv$ "is directly proportional to". It is convenient to use it in cases like this since each of the three statements, if expressed in the usual form, would require a different constant. The symbol is often read "varies . . . with". Thus $y \propto x$ would be read "y varies directly with x," while $y \propto \frac{1}{x}$ reads "y varies inversely with x".

§ 2. *Generalization.*—The result just obtained can easily be generalized. Let P, Q and R be any three variables whose values are at any moment x, y, z. Suppose also that:—

$z \propto x$ when y is constant
$z \propto y$ when x is constant
then shall $z \propto xy$ when both vary.

Let k be the value of R when P and Q both have unit value. Let Q remain of unit value while P's value changes from 1 to x. At the same time let R's value become z'. Then, by the first relation we have $z' = kx$. Now that P has reached its final value let it remain constant and let Q begin to change from

1 to y. Then R which has the value $z' = kx$ when Q begins to change will (by the second relation) reach a value z such that

$$z = z'.y$$
$$= kx.y$$

when Q has reached its final value y. This argument holds good whatever be the final values, x and y, of the variables P and Q. Thus

$$z = k.xy \text{ or } z \propto xy.$$

§ 3. *More Complicated Cases.*—This weight of liquid (W) that can be poured into a cylindrical tin of given cross-sectional area (A) varies directly with the height (h). That is :—

$$W \propto h \quad (A \text{ constant})$$

Similarly $W \propto A \quad (h \text{ constant})$

and, in accordance with § 2, $W \propto Ah$ when both vary.

But we have also $A \propto r^2$ where r is the radius of the base. Hence we may write $W \propto hr^2$.

This result when generalized takes the form :—

If $z \propto x$ (y constant)
and $z \propto y^2$ (x constant)
then $z \propto xy^2$ (if both vary).

The proof will be conducted precisely as in § 2.

If in the above problem we wished to calculate the height of the tin required to hold a given weight of liquid when the base had a given area we should have :—

$$h \propto W \quad (A \text{ constant})$$
$$h \propto 1/A \quad (W \text{ constant})$$

and, when both vary, $h \propto W/A$.

Here we have an example in which a case of direct proportion or variation is combined with a case of inverse proportion or variation. Moreover since $A \propto r^2$ we can transform it into a still more complicated case of " conjoint variation " :—

$$h \propto W \quad (r \text{ constant})$$
$$h \propto 1/r^2 \quad (W \text{ constant})$$
$$h \propto W/r^2 \quad (\text{when both vary}).$$

These results can both be generalized and established in their generalized form as in § 2.

[Ex. XXV may now be worked.]

SECTION II.

DIRECTED NUMBERS.

THE EXERCISES OF SECTION II.

*** The numbers in ordinary type refer to the pages of *Exercises in Algebra*, Part I; the numbers in heavy type to the pages of this book.

EXERCISE		PAGES
XXVII.	The Use of Directed Numbers	155, 181
XXVIII.	Algebraic Addition and Subtraction	161, 184
XXIX.	Directed Products	168, 193
XXX.	Summation of Arithmetic Series	175, 199
XXXI.	Algebraic Multiplication	183, 207
XXXII.	The Index Notation	187, 214
XXXIII.	Negative Indices	191, 217
XXXIV.	Factorization	193, 176
XXXV.	Algebraic Division	196, 222
XXXVI.	Geometric Series	201, 224
XXXVII.	The Complete Number-Scale	208, 228
XXXVIII.	Further Examples on Directed Numbers	213, 230
XXXIX.	Linear Functions	217, 258
XL.	Directed Trigonometrical Ratios	221, 261
XLI.	Surveying Problems	224, 256
XLII.	Hyperbolic and Parabolic Functions	229, 264
XLIII.	Quadratic Equations	235, 270
XLIV.	Further Equations	239, 240
XLV.	Inverse Parabolic Functions (I)	242, 274
XLVI.	Inverse Parabolic Functions (II)	247, 243
XLVII.	Area Functions	250, 279
XLVIII.	Differential Formulæ	255, 282
XLIX.	Gradients	258, 252
L.	The Calculation of π and the Sine-Table	264, 292

CHAPTER XVII.

THE PROGRAMME OF SECTION II (EXS. XXVII-XXXVIII).

§ 1. *The Use of Directed Numbers* (Ex. XXVII).—It was explained in ch. v., § 3, that the distinctive task of Section II is the study of the consequences which follow when algebraic symbols are understood to represent numbers with signs instead of the signless numbers of arithmetic. The object of Ex. XXVII is to give the pupil his first introduction to these new numbers. Simple as the lesson is, it is one of critical importance.

Historians of mathematics have remarked upon the slowness with which ideas about the nature of positive and negative numbers developed into clearness in the early days of algebra. It is, perhaps, not too much to say that the difficulties which the older thinkers found or raised remain, like an undispersed fog, to obscure the path of the beginner to-day. These difficulties all have the same origin—the failure to perceive that the difference between positive and negative is not synonymous with the difference between greater and less. This confusion of two distinct ideas vitiates, for instance, the example with which beginners have made first acquaintance with negative numbers in every generation since the end of the sixteenth century. If a person (we are told) who possesses £5 is to estimate his wealth as "$+5$," then a person who has no money in his purse but *owes* £5 must write *his* wealth down as "-5," for *he has £5 less than nothing.* Now the difficulty which the beginner has to face here lies not in any abstruseness of the new idea but in its absurdity; for he is asked, as the price of admission to the new subject, to give up the conviction of common sense that there cannot be anything less than nothing. So unfortunate a result might well have suggested a careful scrutiny of positive and negative numbers with the object of determining whether it is inevitable,

or whether the true nature of the numbers had been misapprehended. It is, therefore, a rather strange fact that it remained for the mathematical philosophers of the last century to put the theory of these important entities upon a more satisfactory basis.

The upshot of their investigations is the discovery that over a large part of the field of mathematics the fundamental idea is not *magnitude* but *order*. Some of the more salient consequences of this discovery are given in the exercises of Part II, Section IV, to which the reader who is unacquainted with these modern developments may turn for information in the first instance.[1] Their application to the question of negative wealth is easily illustrated. Suppose a company of persons of varied financial position to be gathered together, and let a humorous tyrant insist upon examining their bank pass-books and arranging them in a single row in order of their financial status. We will suppose his edict to be that every man must stand so that his left-hand neighbour is in a better, and his right-hand neighbour in a worse, financial position than himself. Then common sense will itself insist that the man who owes £5 shall be to the right of the man who has nothing and owes nothing, and that a man who owes £6 shall stand on his right and one who owes £4 on his left. Now let the tyrant proceed to label his victims with numbers indicative of their relative status. He may, perhaps, start by assigning to the possessor of £5 the numerical label "5," and, proceeding on this plan, will mark the man who has neither possessions nor debts with the label "0". He has now to deal with the unfortunates who have to a greater or less extent overdrawn their banking accounts. In labelling these he may reasonably adopt the view that the man who owes his banker £5 is, in respect of his financial position, just as far removed in one direction from the person marked zero as the possessor of a balance of £5 is in the other direction. To signify at once the resemblance and the difference in their situations he may adopt the device of prefixing to the common number "5" the distinguishing signs + and −. If these signs are distributed in a similar way to all the participants in this one-sided game, we shall have the following results.

[1] He will find the treatment most suitable for his purpose in Young's *Fundamental Concepts of Algebra and Geometry*.

All persons on one side of the zero-person will bear numbers prefixed with the sign *plus*, the numbers increasing as the eye moves along the row to the right. All on the other side of the zero-person will have numbers prefixed with the sign *minus*, and these numbers will increase towards the left. Thus the numerical label, together with the sign, serves to fix the position of each person in a perfectly definite way, so that they could instantly be rearranged as before if the row were to be temporarily broken. Moreover the device chosen by the tyrant would obviously enable him to assign a definite place and label to an endless number of fresh victims.

This example, in spite of its triviality, brings out the real significance of positive and negative numbers. Their essential function is not to measure magnitude but to register position in a series in which the terms are arranged upon some principle of order. They are most appropriately used when the series is either actually or potentially extensible without end on both sides of any given term, and they serve to indicate the (real or metaphorical) distance and the (real or metaphorical) direction of the various terms from a term selected as the point of reference and marked zero.

The reader will now see why the name "directed numbers" is given in this book to numbers accompanied by *plus* or *minus* signs. Such a number always refers, explicitly or implicitly, to a term of an ordered series, and the sign always indicates the direction in which the term is to be sought, starting from a certain point of reference or zero. In many cases—and those among the most important—the word "direction" is to be understood literally; in other cases it is only a readily intelligible metaphor. But, in every case in which positive and negative numbers are applicable, direction in one of its senses must be present and constitutes the essence of the conditions which make their use appropriate.

From these considerations an important practical conclusion follows. If positive and negative numbers are used only when the idea of direction is present, then it will obviously be best to teach their use first in cases where the element of direction is explicit and has the familiar spatial character. This is the procedure followed, not only in Ex. XXVII, but also in the subsequent exercises in which the class are to explore the laws of manipulation of directed numbers.

The exposition suggested in ch. XVIII., A, needs no further

introduction. Nevertheless it may be well to emphasise the recommendation made in the lesson, that *plus* and *minus* signs should consistently be prefixed to directed numbers.[1] The adoption of this simple custom proves wonderfully effective in maintaining clearness of ideas, and, therefore, well repays the trifling trouble and delay which it involves. In particular it will serve to keep the young algebraist on his guard against the notion—generally harmless enough, but occasionally a source of mischief—that if a number has no sign it is necessarily positive.

§ 2. *The Manipulation of Directed Numbers* (Exs. XXVIII, XXIX).—The next two exercises deal with subjects which, by general consent, are among the most difficult in the teaching of elementary algebra. These are the rules which are to be observed when directed numbers are added and subtracted, multiplied and divided. As was pointed out in ch. v., § 2, the problem of the "rule of signs" is met with at four distinct points in the theory of algebra. It arises first when we are called upon to evaluate such an expression as

$$46 - 3(15 - 8).$$

Here, as we saw in Ex. XIII and ch. IX., B, the problem is an arithmetical one and can be solved by arithmetical considerations. We have merely to take from 46 three times the difference between 15 and 8, that is, three times 7 or 21. But if the problem is to find the value of $a + b$ or $a - b$ when $a = + 8$, $b = - 4$, the case is very different. The root of the difficulty in which we find ourselves is that the signs connecting the a and b can no longer mean "add" and "subtract" in the common meaning of those terms. This is evident from the fact that when we "add" the $- 4$ to the $+ 8$ the arithmetical operation which actually has to be carried out is subtraction, and is addition when we "subtract" it. Thus our first task is to determine what the signs *plus* and *minus* mean when they connect symbols which stand for directed numbers. It is evident that a similar difficulty will face us if we seek to evaluate the forms

$$ab \text{ or } a/b$$

when the numbers represented by the symbols are directed.

[1] Mr. G. St. L. Carson has suggested writing $\bar{5}$ instead of $- 5$, etc., when the *minus* belongs to the number. The one objection to this proposal is that it would be inconvenient, at any rate in print, to deal similarly with a *plus*.

For what can possibly be the meaning of an instruction to take $(+8)(-4)$ *times ?* Thus here again the first task is to interpret a familiar piece of symbolism in circumstances in which it no longer bears its usual meaning.

It is worth while to remark that the difficulties here in view affect the manipulation of positive equally with negative numbers. It is generally supposed, for example, that although the determination of the value of the "product" of -4 by -8 constitutes a real problem, the evaluation of $(+4) \times (+8)$ does not. It is assumed that the answer in the latter case is obviously $+32$. But consideration will show that this easy triumph is a hollow one, being gained by the totally illegitimate identification of $+4$ and $+8$ with the signless numbers 4 and 8. The truth is that, without some definite convention or definition, "multiplication" in the case of two positive numbers is just as mysterious, and, indeed, meaningless an operation as in the case of two negatives. Both cases are equally in need of examination.

There are two distinct ways in which these problems may be attacked. The first is by the "high *priori* road" of arbitrary definition and postulate. We may, for example, say that the operation denoted by the symbol $\times$ is to mean ordinary multiplication when it connects two non-directed numbers, and when it connects directed numbers is to be taken as *defined* by the following laws :—

$$(+p) \times (+q) = +pq, \qquad (-p) \times (-q) = +pq$$
$$(+p) \times (-q) = -pq, \qquad (-p) \times (+q) = -pq$$

Of this method it is sufficient to say here that although theoretically unimpeachable it is utterly unsuitable as an introduction to the subject. It is impossible for the young student to appreciate or to profit by a mode of treatment which presupposes a trained power of logical detachment and abstract reasoning. In practice it can lead only to "symbol juggling" of the most unsatisfactory kind, in which the student carries out manipulations in accordance with given rules but has only the most shadowy idea of their real meaning.

The general principle that all forms of algebraic symbolism should be taught as natural ways of describing certain facts suggests a method of dealing with this question much more suitable to the needs of the beginner. The essence of the method is the rule that *the results of a given manipulation*

of the algebraic symbols shall always accord with the known behaviour of the things to which they correspond. Thus the formula $d = vt$ is an obvious way of describing the position at any moment of a point moving with uniform velocity v. Moreover the three elements to which the three symbols correspond are all capable of description by directed numbers. Thus the study of the relation between this formula and the behaviour which it is intended to describe gives a good opportunity for determining the rules to be followed in evaluating a piece of symbolism of the form vt when the symbols represent directed numbers. Now it is easily shown (as in ch. XVIII., C) that when both the velocity of the point and the moment of time under consideration are to be denoted by negative numbers then the distance of the point from its zero will always be denoted by a positive number. From this known *fact* about the things referred to we draw a definite rule for the manipulation of the symbols which refer to them ; namely, that the product of two negative numbers must always be considered to be positive.

The reader may object that the conclusion goes beyond the evidence, and that this rule for the product of two negative numbers has been proved to hold only when they represent respectively a velocity and a time. This objection is removed by the argument of ch. XVIII., C, which shows that the rule must hold good for the products of all pairs of numbers which can be represented graphically upon the same principle as velocity and time in the case of a moving point. The reader may readily satisfy himself that this condition covers all cases in which the ordinary laws of algebraic multiplication apply, and that it fails to cover only those cases, such as the product of vectors, which as a matter of fact lie outside the scope of the ordinary rules of algebra.

§ 3. *Algebraic Addition and Subtraction* (Ex. XXVIII, ch. XVIII., B).—Before products and quotients of directed numbers are considered the case of sums and differences is discussed in ch. XVIII., B. We have said that the root difficulty here is that in speaking of a "sum" or "difference" of directed numbers we cannot be using these words in their ordinary arithmetical sense, and that we have, therefore, to make clear the new meanings which are to be attributed to the familiar terms. But closer examination shows that this statement does not express very accurately the nature of our

task ; it will be found that we have not so much to discover new meanings for the words "sum" and "difference" as to determine what parts of their common meaning do and what parts do not apply to directed numbers. Perhaps the best way to discover the common element is to examine a case of "addition" in which the addenda are not numbers at all. A simple case of this kind is the addition of "vectors" illustrated by fig. 21, p. 112. In this figure the lines AC and CB represent two vectors and AB represents a third vector which is called their "sum". Thus if the three vectors are denoted in order by the letters a, b and c, the relation between them is expressed by the formula

$$c = a + b.$$

Now what is the meaning of this formula? Remembering that the vectors represent certain movements whose directions are parallel to those of the lines and whose magnitudes are represented to scale, we see that the formula is equivalent to the statement: If you take the movement a and take also the movement b your ultimate position will be the same as if you had taken the movement c. It is clear that the one element common to the use here of the sign "+" and to its use in arithmetic is the idea represented by the words "and take also". If I take 4 books "and take also" 6 books the result is the same as if I had taken 10 books. This is the meaning of the statement that $4 + 6 = 10$, or that 10 is the "sum" of 4 and 6. Thus, if some one asks me what is the "sum" of 4 books and 6 books and also what is the "sum" of a movement of 4 yards in one direction and a movement of 6 yards in another direction, I recognize that the element of identity in the two problems which justifies the use of the same word in enunciating them is the fact that I have to consider in each case the result of *taking two components together*. But the identity between the problems stops at this point. To find the resultant (as it may be called) of the two numbers of books I have to follow one rule—the rule of arithmetical addition. To find the resultant of the two movements I have to follow another rule—the rule of "vector addition".

Similar considerations bring out the common meaning contained in different uses of the terms "subtraction" and "difference". Thus, in fig. 21, if the vectors b and c are given and the vector a is required, the process by which a is

determined is called "subtracting" the vector b from the vector c and is represented by the formula

$$a = c - b$$

To understand how this process can be called subtraction we must note that c is here regarded as the resultant of two components, a and b, of which b is known and a unknown. To find the unknown a we draw the combination $a + b$ which is equivalent to c and then discard the b. Thus we have

$$a = (a + b) - b$$
$$= c - b$$

as above. Similarly, if I want to know the result of subtracting 4 from 10 I remember that a whole composed of 10 books, or other countable objects, can be regarded as made up of two parts composed of 4 and 6 respectively. Hence if I remove or ignore the 4 I am left with 6 as the result of my subtraction. Thus the common element in all problems of subtraction is that we are given a resultant of two components and also one of the components, and are required to determine the other component. But, as in the case of addition so here, rules for making this determination in any given case will depend upon the known properties of the things combined.

The foregoing considerations underlie the treatment of algebraic addition and subtraction in ch. XVIII., C. The argument is, in brief, as follows. A piece of symbolism such as

$$-6 + 2 = -4$$

may represent the solution of two distinct problems. In one of those problems we may have been told that two movements of -6 and $+2$ have been made in succession, and the question is what is the single movement to which they are equivalent. In the other we may have been told that two successive movements have resulted in a displacement of -6, one of these movements being -2, and the question is what was the other component movement. Now our uncertainty as to what the problem was comes from the fact that the numbers are directed. Similar problems involving numbers without signs could be discriminated at once: if the numbers were connected by a *plus* the problem would be one of the first kind, if by a *minus* one of the second kind. The question arises, therefore, whether the signs *plus* and *minus* cannot be used to discriminate the problems symbolically when the numbers are directed as well as when they are undirected.

The result of this inquiry is the decision that the first problem may be written in the form

$$(-6) + (+2) = -4$$

and the second in the form

$$(-6) - (-2) = -4$$

The result of this part of the discussion is, therefore, to bring out the fact that, in pieces of symbolism like those just quoted, the signs + and − are each used for two distinct purposes, neither of which is the same as its ordinary arithmetical use. The signs within the brackets are attached to, and form part of, the numbers, and are indicative of direction. The signs between the brackets have a totally different function—they are there simply to show to which of two types the given problem belongs.

The first stage in the actual solution must be the removal of these latter signs, which have done their work when they have informed us of the nature of the problem under consideration. The rules for manipulating the signs attached to the numbers when the enclosing brackets are removed form a very important detail of the argument. The method followed is the one explained in § 2. By considering the actual and known results of typical problems of each kind we determine what rules of signs must be adopted in order that the manipulations of our formulæ may agree with the known behaviour of the things to which they refer.

Ex. XXVIII, A, consists of simple examples intended to illustrate and to "drive home" the distinctions and the rules reached in the course of the lesson. **No. 1** may be taken orally. **Nos. 2** to **9** have a flavour of the puzzle about them which makes them surprisingly attractive to the young student. In division B the newly acquired ideas are applied to the making of formulæ in which the symbols are to stand for directed numbers. These formulæ illustrate the striking gain in generalizing power which comes with the adoption of these modes of expression. Cases which formerly could be exhausted only by using several different formulæ are now found to come within the scope of a single one. After working the first few of the examples the student should be convinced that to obtain a formula applicable to all possible cases it is sufficient to consider a case in which the numbers either are non-directed or are all positive. The formula which describes this case will be equally applicable to all others. The ex-

amples of division C may be considered as optional. They are inserted rather as a tribute to convention than because they are of any particular use at the present stage of the pupil's progress. But if done at all they should certainly be done with the careful attention to their proper significance which is illustrated in ch. XVIII., B, §§ 6, 7.

§ 4. *Products and Quotients of Directed Numbers* (Ex. XXIX ; ch. XVIII., C).—The method followed in dealing with the rules of signs of products and quotients of directed numbers has been sufficiently explained in § 2. In Ex. XXIX, A, the student is shown how these rules make it possible to widen still further the range of cases included in the reference of a single formula. Thus these examples supplement those of Ex. XXVIII, B, and complete our study of the primary consequences of permitting the symbols of a formula to represent directed instead of non-directed numbers.

The examples of division B illustrate the power of the new methods of algebraic statement by applying them to the description of the terms of an arithmetic sequence—that is, an endless series of numbers exhibiting from term to term a constant difference. The objections which good authorities have urged against the conventional treatment of arithmetic progressions would probably not be maintained here. The concept of the doubly-endless sequence of directed numbers is not only most attractive to young students but is also of fundamental importance throughout algebra. If a practical justification for its introduction is required we may point to the theory of logarithms in which the central idea (at least as it is treated in this book) is the correlation, term by term, of a doubly-endless arithmetic sequence with a doubly-endless geometric sequence. For the present we rely upon its intrinsic interest and theoretical importance and its value as a means of demonstrating the powers of directed numbers.

In division C the examples deal with "progressions," that is with sections of a complete sequence counting forwards from a given term. Here again the subject is introduced episodically and as a further study in the properties of directed numbers, and not as a new "rule". The examples have, however, special importance in connexion with the summations which are to be studied in the next exercise.

§ 5. *The Summation of Constant-difference Series* (Ex. XXX, A ; ch. XIX., A).—The summation of arithmetic series

is a rather trivial occupation regarded as an end in itself. The indubitable attraction which it has for the young mathematician may, however, as in the exposition of ch. XIX., A, be made a convenient means of introducing a concept—that of positive and negative areas—which will be found of great use in developing the theory of algebraic multiplication in ch. XX. This consideration alone would suffice to justify the attention given to the subject. But as will shortly be seen, this process of summation is preliminary to an enterprise of considerable moment—namely, the student's first excursion into the province of the integral calculus. Thus it has much practical as well as theoretical significance.

The exposition of ch. XIX., A, needs no supplementary explanation. The teacher is advised to adopt the device there described, and to distinguish between positive and negative areas by the use either of coloured paper or of coloured chalk. The pupils should use the same device in their answers to Ex. XXX, No. 1. Time may be saved by distributing long strips of red and blue paper of a convenient width which have previously been coated with gum upon the reverse side. From these it is easy to cut columns of the required length to be stuck in the exercise book.

§ 6. *The Calculation of certain Areas and Volumes* (Ex. XXX, B ; ch. XIX., B).—The grounds for including an elementary treatment of the ideas of the calculus in the algebra course have been stated in ch. II., § 3, where the general spirit in which it should be undertaken was also discussed. We have now reached the point at which the systematic development of the ideas there defended is to begin. It will be observed that we start, not with differentiation, but with integration. Many reasons may be given for this order of procedure, but the teacher whose class has worked Ex. XXX will probably agree that the simplicity of the argument and the power which it puts at once into his pupils' hands are a sufficient justification even for so radical a departure from tradition. If a further defence is required we point to the fact that the exposition of ch. XIX., B, follows closely the track which led our famous countryman, John Wallis (1616-1703), to what was practically the invention of the integral calculus. Wallis built, of course, upon the ideas of his predecessors—notably Archimedes among the ancients and Cavalieri (1598-1647) among the moderns—but he was the first mathematician

to attempt a systematic " arithmetic of infinites ". His work was the starting-point of Newton, Leibniz, and the other mathematicians who gradually gave to the calculus the shape in which we know it and use it to-day. But, as in so many cases, the later form of the science, although technically more perfect and a more effective instrument of mathematical investigation, is not so suitable for the purposes of the teacher as the earlier form. The beginner gets a more rapid insight into the meaning of the new ideas and a more complete mastery over them if they are presented to him as they occurred to the pioneer and in connexion with attempts to solve the same kind of problems. For this reason the method and subject-matter of ch. XIX., B, are borrowed, almost as they stand, from the first chapter of Wallis's *Arithmetica Infinitorum* (Latin folio edition, Oxford, 1699).

The teacher will find it useful to prepare for exhibition to his class the models which correspond to figs. 54 and 55. It is convenient to have in duplicate the one represented in fig. 55 and to fix the grey rectangles of one of the two specimens in the positions shown in fig. 56. It is interesting to note that the argument used in connexion with these figures and *Exercises*, fig. 34, is taken from Cavalieri, *Geometria Indivisibilibus Continuorum*, a work from which Wallis admits drawing much inspiration.

The examples of Ex. XXX, B, are of great importance and should be worked very thoroughly, for they illustrate in a simple and concrete manner the central problem of integration and the way in which it may be solved. Briefly, the problem is : Given the law followed by the rate of change of a thing's magnitude, what is the law governing the magnitudes themselves? The contribution of the present examples to the solution of this general problem is that if the rate of change (v) of the magnitude of a variable x follows the law

$$v = a + bt$$

then the magnitude of x is itself given by the formula

$$x = at + \tfrac{1}{2}bt^2$$

The examples contain various applications of this interesting result, but the most important are those in which the variable is a distance and the rate of change of its magnitude the velocity of a moving point. These examples are not to be regarded as a trespass into the province of mechanics. It is

true that a text-book of mechanics always contains a section in which the formulæ $v = at$, $s = \frac{1}{2}at^2$, etc., are proved and applied to kinematical problems, but this circumstance is to be taken rather as evidence that the teacher of mathematics leaves part of his proper work undone than as an argument against the inclusion of these same problems in this book. The business of the student of mechanics is to determine what forms of motion are actually exhibited by material bodies, to define them quantitatively, and to determine the physical conditions which govern their appearance. The investigation of the mathematical consequences of laws of motion suggested by common observation and experience is no more an invasion of his territory than is the investigation of the trigonometrical relations of which he also constantly makes use. Motion is simply "geometry *plus* time," and any reason which justifies the study of geometry as a branch of mathematics must justify equally the inclusion of kinematics.

In **Nos. 10-14** it is important to note that the constancy in the final *rate* in question is represented in the graph by the constant *height* of the last column. This is true however narrow the column may be—that is, however short a time is under consideration. Thus when the increase in the rate is continuous the rate at any moment is measured by the height of the corresponding ordinate. This argument gives the answer to No. 14. It also prepares the pupil for the consideration of a rate as a "limit"—an idea which will demand careful attention at a later stage of our work (see ch. LI.).

In solving problems (such as Nos. 19, 21, 22) in which the units employed in measuring the rate are not identical with the unit represented along the horizontal axis the teacher is strongly advised not to deal with the "change of unit" by any formula or similar manipulation of the numbers, but to work directly on the principle that the area of the diagram represents the space covered in the given time. Thus in **No. 19** the abscissæ measure minutes and the ordinates miles per hour. The triangle whose base is 12 and height 60 has an area of $\frac{1}{2}$ of 60×12, that is 360 units, and represents the distance travelled in 12 minutes, which will be 6 miles. Thus each unit of area of the figure represents a distance of 6/360, or 1/60 of a mile. (Note that the unit of area will not be a square unless the horizontal and vertical scales of the diagram are identical; in general it will be a rectangle.) Now

the law for the velocity in miles per hour measured after t minutes is evidently

$$v = 5t$$

so that the area of the triangle whose base is t will be $\frac{1}{2}$ of $5t \times t$. Applying the condition that each unit of area represents 1/60 of a mile we have the formula

$$s = \tfrac{1}{2} \text{ of } 5t \times t \times \tfrac{1}{60}$$
$$= \tfrac{1}{24}t^2$$

No. 29 represents the type of calculation in which the argument reaches its culmination. Given the relation

$$v = 7{\cdot}2 + 6{\cdot}4t$$

the student should now be able to proceed at once to the conclusion

$$s = 7{\cdot}2t + 3{\cdot}2t^2$$

The argument here differs from an integration only in the absence of the traditional notation in which integration-processes are commonly expressed.

§ 7. *Algebraic Multiplication* (Ex. XXXI, ch. xx.).—The problem studied in ch. xviii., C, was to find the single directed number represented by the symbolism ab when both a and b represent given directed numbers. Thus it is a problem which concerns not the manipulation of a formula but its arithmetical evaluation. The problem of algebraic multiplication is quite different from this, for it concerns the validity of algebraic identities when the factors of one of the equivalent expressions are themselves algebraic sums or differences. Consider, for example, the familiar identity

$$(a + b)(a - b) = a^2 - b^2$$

When a and b are symbols for signless numbers and + and − mean "add" and "subtract" in the ordinary senses of the words there is no difficulty in proving the validity of this general statement by the process which we learnt in ch. vii. to call "multiplication". But when a and b become symbols for directed numbers + and − lose their ordinary significance and acquire the new meanings explained in ch. xviii., B, and in § 3 of this chapter. It would, therefore, be quite illegitimate to assume that the argument which established the equivalence in the former case suffices to establish it also in the new. It is evident that the new case must be examined upon its own merits.

The task of ch. xx., A, is to prove that the method of

algebraic multiplication which may be used to find and to prove identities, when the symbols stand for non-directed numbers and the connecting signs imply arithmetical addition and subtraction, does as a matter of fact hold good when the symbols stand for directed numbers and the connecting signs imply algebraic multiplication and subtraction. When this point has been established we shall have proved that the symbolism which describes the relations and manipulations of non-directed numbers holds good in every respect for directed numbers. Thus any result which has been shown to be true for the one kind of numbers may henceforward be assumed to be true also for the other kind. In other words we shall have proved that a single rule governs the combinations of *plus* and *minus* signs over the whole range of their significance.

A word may be added about the method of investigation adopted in ch. xx., A. It would, no doubt, be possible to find a general method which would prove at once that the rule of signs must hold good when *plus* and *minus* signify algebraic addition and subtraction. A general proof has, however, the disadvantage that it does not bring the student's mind into that close relation with particular cases which is the necessary foundation of full and exact knowledge. For this reason the opposite plan is adopted and the student is led to establish his hypothesis by examining all the standard cases. It will also be observed that the method of examination followed in ch. xviii. is again followed here. That is, the rule which governs a particular kind of manipulation is determined by observing the actual behaviour of the things which the formulæ describe. The details of the lesson—including the use of the coloured rectangles—are given too fully to need further elucidation. It will be noted that one set of coloured diagrams is given as the frontispiece of *Exercises*, Part I, in order that the pupil may have it before him in working Ex. XXXI, A.

Of the examples in divisions B and C it is sufficient to say that they are chosen to illustrate the real significance of algebraic multiplication—namely, that it is a process by which we can predict the typical form of a product when we know the typical forms of its factors. The danger in this part of algebra is always that the pupil may forget that his results are valuable only for their analytical significance, and may

come to suppose that the power of manipulating symbols skilfully is itself to be counted as a virtue.

The fact that the process of algebraic multiplication is always the determination of an identity is best kept before the student's mind by insistence upon the mode of setting down illustrated in a simple case in ch. xx., A, 5, and previously in ch. vii., B, § 2. Thus Ex. XXXI, **No. 32** (ii) should be worked as follows :—

$$\begin{aligned}(3t^2-2t+1)(t^2-2t+3) &= 3t^4-2t^3+\ \ t^2\\ &\quad -6t^3+\ 4t^2-2t\\ &\quad +\ 9t^2-6t+3\\ &= 3t^4-8t^3+14t^2-8t+3\end{aligned}$$

One of the advantages of postponing algebraic multiplication to the point here assigned to it is the possibility of illustrating the power of the process by setting the pupil to solve simple problems of the type which will be recognized later as covered by the formula of the binomial theorem. A selection of these is given in Ex. XXXI, D. The coefficients of the various expansions are to be determined by the empirical method which was the only one known until Newton discovered the general rule by which any binomial expansion can be written down without reference to the coefficients of any other. This generalization is reserved to Ex. LXVIII.[1]

§ 8. *The Theory of Integral Indices* (Exs. XXXII, XXXIII ; ch. xxi., A, B).—One of the first algebraic idioms which the pupil learnt was the use of the index to symbolize the square, cube and occasionally the fourth power of a variable. Many of the manipulative processes studied—in particular those of the last exercise—have also involved raising the index of a variable by multiplication. Nevertheless there has been no point at which there was anything to be gained by a systematic examination of the principles which underlie the use of the integral power-index, and the pupil has been left to deal with the particular cases as they arose without being called upon to formulate the rules of his procedure. The point has now been reached at which such a formulation would be useful ; it is made, therefore, the subject of ch. xxi., A, and is illustrated by Ex. XXXII. In accordance with our

[1] The author is indebted to Mr. C. O. Tuckey of Charterhouse for the idea of introducing at this point some preliminary exercises in binomial expansions.

usual plan, the theoretical question of the "laws of indices" is raised in connexion with a practical problem—namely the problem of representing concisely and manipulating conveniently the large numbers which so frequently occur in physics and in the statistics of demography and economics.

The indices of ch. XXI., A, and Ex. XXXII are non-directed or signless numbers; ch. XXI., B, and Ex. XXXIII introduce the highly important idea of indices as directed numbers. Once more the argument has its origin in practical considerations. It is shown that the representation and manipulation of small numbers would be facilitated if we had a method of symbolizing repeated division similar to the index method of symbolizing repeated multiplication. The principle of continuity suggests the negative index as the most convenient solution of this problem. There then arise the questions: (i) what rules must be followed when negative indices are to be combined with one another, and (ii) what rules must be followed when negative are combined with positive indices. The inquiry takes the course made familiar by our investigations of the rules of signs, and leads to the conclusion that the laws of indices established for non-directed indices also hold good when the indices are directed.

Attention may be called to a detail of the method. In both divisions of the chapter the index is studied not so much as a notation for expressing numbers as for expressing operations upon numbers. Thus the question raised does not directly concern the meaning of the symbolism a^n but that of the symbolism $\times a^n$. The distinction is not one of vital importance, but the adoption of the point of view here recommended will certainly be found to add clearness and interest to the discussion. As the result of it the student should possess the perfectly simple and definite notion that $\times a^{+p}$ means p successive multiplications by the number a, and $\times a^{-p}$ p successive divisions. The *number* a^{+p} can be regarded as equivalent to either $1 \times a^{+p}$ or $a \times a^{+p-1}$, and the number a^{-p} as equivalent to either $1 \div a^{+p}$ or $a \div a^{+p+1}$. It is supposed here, of course, that p is non-directed.

The term "standard form" used in Ex. XXXIII, though becoming widespread, cannot, perhaps, be regarded as universally known. A number is expressed in the standard form when it is written as a decimal with a single digit before the decimal point multiplied by some positive or negative power

of ten. Thus the standard forms of the numbers 18574 and 0·0005937 are respectively $1{\cdot}8574 \times 10^4$ and $5{\cdot}937 \times 10^{-4}$. Facility in expressing numbers in the standard form will be found very serviceable in dealing with logarithms.

It is usual to teach fractional positive indices before negative indices are introduced at all. The teacher may possibly think that the simplicity and directness of the exposition suggested in ch. XXI., B, justify a departure from this custom; nevertheless, it may be worth while to add a brief note upon the point. In the first place it should be observed that until the pupil comes to the study either of logarithms or of "Wallis's law" of integration he has no occasion to feel the need of fractional indices; upon the general principle which should govern an elementary treatment of the subject there is, therefore, no justification for introducing them. In the second place it is obvious that the step from integral to fractional indices is, from the theoretical standpoint, much greater than the step from positive to negative integral indices. As we have seen, nothing is involved in the latter step except the extension of the scope of the symbol $\times a^n$ to include successive division as well as successive multiplication. But, as the argument of ch. XXXIV. shows, the interpretation of a^n, when n is fractional, involves considerations so much more complicated than those of the present chapter that they should certainly be left until a later stage. Many weighty authorities would exclude them altogether from the elementary course; but if included there they should unquestionably be taken in connexion with the theory of logarithms (cf. ch. V., 6).

§ 9. *Algebraic Division. Constant-ratio Series* (Exs. XXXIV-XXXVI, C; ch. XXII.).—The process of algebraic division is, of course, the inverse of algebraic multiplication. There are, however, two ways of regarding its results sufficiently different to demand separate attention. According to the first the problem of division is simply to retrace the steps taken in multiplication, that is, to find the two factors of a given product, or, given one factor, to find the other. This point of view is represented in the examples of Ex. XXXIV, including the special case—not of any particular practical importance—in which the two factors are identical (division B). The second point of view is presented in Ex. XXXV. Here the typical problem is to find an integral expression

equivalent to a given algebraic fraction, or, if no such exact equivalent exists, to find an equivalent consisting of an integral expression together with the simplest possible "fractional complement".

The practical importance of this process appears in division C where it is employed to yield "expansions" of fractional expressions, such as $1/(1-a)$, which may be carried as far as any desired power of the variable. When $|a|<1$[1] these expansions can also be regarded as "approximation-formulæ" representing the values of the fractions to any desired degree of accuracy. The reasoning upon which this statement is based deserves careful attention in view of its application in the theory of endless geometric, or constant-ratio, series. After a given number of the terms of the expansion have been computed the difference between their sum and the real value of the fraction is given exactly by the fractional complement. Thus the complement is always an exact measure of the degree of error involved in the adoption of a given approximation-formula. Moreover, by taking a sufficient number of terms it seems evident that the complement (and therefore the error) can be made smaller than any number that can be named. An exhaustive analysis of this conclusion need not necessarily be taken at this stage. The student should, however, see that it involves the assumption that if the product of a number a by a factor r produces a number b which is less than a, then a sufficient number of repetitions of the multiplication will reduce the product below any given value however small. On the other hand it may prove desirable to submit this assumption to further examination in order to make it more acceptable to the critically-minded student who occasionally appears in our classes even at the age of thirteen or fourteen. If l is a non-directed number >1 it is certain that $1/l$ is <1 and that by making l large enough $1/l$ may be made less than any number that can be named. Now let any non-directed number a be multiplied by a number $r=1+i$, both r and i being non-directed, and let the product be b. Then it is clear that $b>a$, for it may be written as $a+ia$. Similarly, when $b=ar$ is multiplied by $r=1+i$, the product $ar^2=b+ib$ must be greater than b. Moreover the excess, ib,

[1] The symbol $|a|$ means "the non-directed number contained in a".

is greater than the former excess ia, for b is greater than a. The argument may evidently be repeated in connexion with each of an endless number of successive multiplications by r. Let l be the number reached as the result of n such multiplications—that is, let $l = ar^n$. Then we have evidently proved that l could also be regarded as reached from a by a series of n constantly increasing steps of which the smallest is the first, ia. Now every one will admit that, however small ia may be, a sufficient number of steps of that magnitude will carry us from a past any other finite number however large.[1] If the steps (with the exception of the first) are all larger than ia the only difference will be that fewer will be required. Thus it becomes certain that when r is >1 we may, by taking n sufficiently large, make $ar^n = l$ larger than any number that any one may choose to name. It follows by the preceding argument that $1/ar^n$ may be made smaller than any number that can be named. But since this conclusion is quite independent of the value of a it is equivalent to the conclusion that, when $r<1$, ar^n may, by taking a sufficiently large value of n, be made as small as we please. This was the assumption to be justified.

The argument underlying Ex. XXXV, C, enables us, in ch. XXII. and Ex. XXXVI, to proceed to summations of constant-ratio series without resorting to the undesirable course of teaching them as a separate "rule". (Compare the treatment of constant-difference series in Exs. XXIX and XXX.) The practical problems presented in Ex. XXXVI are to be regarded simply as cases for the immediate application of the equivalences established in Ex. XXXV, Nos. 24-6. In every case it is important to realize the significance of the "complementary fraction" which measures the degree of approximation of the calculation. This is well brought out by the graphic exercises included in some of the examples. It would probably be inconvenient to avoid altogether the traditional term "sum to infinity" but the teacher should see that no mystical idea is allowed to attach to it. The so-called "sum to infinity" is not the sum of any number of terms at all; it is merely the value of the series $a(1 + r + r^2 + r^3 + r^4 + \ldots)$ increased by the value of the appropriate fractional complement. Its usefulness consists simply in the fact that

[1] This assumption is known as the axiom of Archimedes.

the sum of any actual number of terms of the series constantly approaches nearer to it as the number increases, and may be made to become *and subsequently to remain* as near to it as we please.

The examples on annuities, etc., with which the exercise closes are an important justification for the introduction of constant-ratio series but hardly require special comment. They are further developed in the later exercises of Section III and are made, in Section IV (Part II), to lead to an elementary theory of life insurance and other actuarial topics. Finally it may be observed that the whole treatment of constant-ratio series is a natural preparation for, and reaches its culmination in, the study of logarithms (cf. § 4 of this chapter).

§ 10. *The Complete Number-scale* (Exs. XXXVII, XXXVIII; ch. XXIII.).—In the preceding exercises of this section the student has examined the consequences of supposing his symbols to represent directed numbers in the case of all the fundamental processes of algebra except two—the simplification of fractions and the manipulations used in changing the subject of a formula. The consideration of these two groups of processes is undertaken in ch. XXIII. and illustrated in the last two exercises of the present group.

The essence of the treatment of these topics in Section I consisted in studying the ways in which we actually manipulate numbers in arithmetic and in basing on those ways our rules for manipulating the symbols of numbers. Thus if we are told that a certain number, less 6, is 12 we know that the number in question must be $12+6$ or 18. Since it is clear that we should proceed to deal with any other similar case in the same way we have the rule that, given

$$n-b=a$$

we may at once write

$$n=a+b$$

But it is evident that this rule has been proved only for the symbols of non-directed numbers, and that (as in the case of identities) a new investigation is necessary before we may use it when our symbols stand for directed numbers.

As a preliminary to this investigation it is suggested in ch. XXIII. that the teacher shall review the properties of directed numbers, making explicit the various matters that have been implicitly assumed at earlier stages of the work. Thus the

pupil is taught to see that the dominant notion underlying the scheme of directed numbers is the notion of *order*, and that the rules for changing the subject of a formula or simplifying fractions must be established by considering not what happens when numbers are added, subtracted, etc., as in arithmetic, but what happens when a point moves backwards and forwards along an endless line representative of the complete scale of positive and negative numbers.

The examples of Ex. XXXVII are all important but contain no technical difficulties. The positions of the points P and Q on the parallel scales of fig. 63 (p. 230) illustrate the following successive stages in the solution of **No. 11.**

$\frac{1}{3}(2x+17)-7=10-\frac{1}{2}(1-3x)$ . . $(OP=OQ=-10)$

$\frac{1}{3}(2x+17)+\frac{1}{2}(1-3x)=17$ . . $(OP_1=OQ_1=+17)$

$\frac{2}{3}x-\frac{3}{2}x=+17-\frac{17}{3}-\frac{1}{2}$

$-\frac{5}{6}x=+10\frac{5}{6}$ . . . $(OP_2=OQ_2=+10\frac{5}{6})$

$x=-13$ . . . $(OP_3=OQ_3=-13)$

The teacher should note that the technical term "equation" is introduced for the first time in connexion with these examples. No doubt it would be inconvenient never to apply it except when a relation is presented in the form $f(x)=0$ but there is much to be said for restricting it, as a rule, to such cases.

Ex. XXXVIII is the last of the first group of Section II. The examples present rather greater technical difficulty than those of the earlier exercises but do not involve any new principles. **Nos. 9 and 10** will be referred to in Ex. XLVII, and **Nos. 11-16** are important as foreshadowing the application of the binomial theorem to the case of a negative integral exponent.

CHAPTER XVIII.

DIRECTED NUMBERS.

A. The Use of Directed Numbers[1] (ch. XVII., § 1; Ex. XXVII).

§ 1. *The Minus Sign as an Index of Direction.*—A boy comes out of a class-room at O having been dispatched on an errand upstairs. When he has reached the 14th stair he drops his pencil and has to go down 5 stairs to pick it up. He is now 9 stairs above the landing. You did not need, of course, to count the steps to find this result. You merely performed in your head the calculation which we can write down thus :—

$$\text{Number of stairs from O} = 14 - 5 = 9$$

We should generally say that the *minus* sign here shows that the 5 has to be taken away from the 14. It is clear, however, that the 14 may be regarded as indicating the distance the boy goes upwards, and the 5 the distance he goes downwards. Thus the *minus* sign is in this problem an indication not only of subtraction but also of reversal of direction. This argument may be repeated with other numbers less than 14, including finally the case in which the pencil rolls on to the landing. In this case $s = 14 - 14 = 0$; the boy has gone up 14 steps and come down as many.

But if the pencil rolls down the upper flight, across the landing and stops only on the 8th stair below, then the boy must descend 22 stairs to reach it. In other words, if he first ascends 14 stairs and then descends 22 he will eventually be 8 stairs below the starting-point. Can a subtraction

[1] A blackboard diagram is required representing a staircase with a landing in the middle marked "O".

sum, arranged upon the same plan as before, give this result? Setting down first the movement upstairs and then the movement downstairs we should have:—

$$s = 14 - 22.$$

Now it is clear that if *minus* is taken to mean "subtract" here the operation is impossible; you cannot take 22 from 14. But if the *minus* means that the boy takes 22 stairs downwards after taking 14 upwards, there is, of course, nothing absurd in the expression. The only question is, can it show us that after his second movement the boy will be 8 stairs below O just as the former calculation showed us that he was 9 stairs above O? Let us write it down again thus:—

$$\begin{aligned} s &= 14 - 22 \\ &= 14 - 14 - 8 \\ &= -8 \end{aligned}$$

If -22 means "Take 22 steps downwards" we may, of course, break it up into two stages: "Take 14 steps downwards on to the landing and then 8 more on the lower flight". These instructions are represented by $-14-8$. But we *can* take 14 away from 14, so the part of the expression which reads "$14 - 14$" can be dropped out altogether. We are left with the expression "-8" as the answer to our problem, and that answer can evidently be taken to mean "8 stairs below O".

§ 2. *Positive and Negative Numbers.*—We now see that when the numbers in a calculation measure distances up or down a scale the sign *minus* can be taken to mean "Move downwards" as well as "Subtract," and *the former meaning remains even when the latter becomes impossible.* That is, we can obtain an answer to our problem when the number with the *minus* in front of it is greater than the other just as well as when it is smaller. We have seen, further, that the answer in the former case will be a number with a *minus* before it, and that this kind of answer always means a position below the "origin" instead of above it.

In such cases we can think of the *minus* sign as attached to the numbers that represent movements or measurements downwards. Thus -5 will be a number which represents a movement downwards on our scale, the *minus* being regarded as a part of the number, just as the 5 is. We can also attach *minus* signs to the numbers on our

scale below the origin to distinguish them from the numbers attached to distances above the origin. Numbers which in this way indicate a movement downwards or a position below the origin may be called *negative* numbers.

If our boy ran up two flights of stairs, the first containing 24 and the second 17 steps, we should find his position at the end by the calculation :—

$$\begin{aligned} s &= 24 + 17 \\ &= 41 \end{aligned}$$

It is clear that just as the *minus* in the former problems could be taken to mean "Go down" as well as "Subtract," so in this problem the *plus* sign can be taken to mean not only "Add" but also "Go up". The *plus* sign is, then, a sign that may be attached to numbers to indicate upward movement or measurement.

In the former cases the fact that the final position of the boy was below the origin was shown by the *minus* in front of the result of our calculation. It will be well to put the *plus* sign in front of the result here to show that the final position of the boy is above the origin, as well as in front of the 24 to show that this number also represents a movement upwards. Thus our calculation will now read :—

$$\begin{aligned} s &= +24 + 17 \\ &= +41 \end{aligned}$$

Similarly it will be convenient to attach *plus* signs to the numbers in our scale that indicate positions above the origin. Numbers which have a *plus* sign attached to them to indicate upward movement or measurement may be called *positive* numbers. Positive and negative numbers may conveniently be called *directed* numbers.

Directed numbers may be used in calculations concerning many other movements and measurements besides those which are taken up or down : for example, movements to right or left, the forward and backward movements of watch-hands, the thermometer scale, etc. The arbitrary nature of the choice of the positive direction should be brought out. In the illustrative calculations *every* directed number should have its sign attached as in the last example above.

§ 3. *Evaluation of Combinations of Directed Numbers.*— If a boy, starting as before from the landing, goes down 6 stairs, then is called back and goes up 11 stairs on to the

upper flight, and finally goes down 18 stairs from that point, where is he at the end of his movements?

It is clear that he will be at the same place as if he had taken the two downward movements one after the other and then the upward movement of 11 steps, or had taken first the upward movement and then the two downward ones in succession. That is:—

$$\begin{aligned} s &= -6 + 11 - 18 \\ &= -6 - 18 + 11 \\ &= -24 + 11 \\ &= -13 \end{aligned}$$

That is, he will be 13 steps below the origin.

It seems, then, that we can alter as we please the order of the numbers representing the movements without altering the result. But it is easier to calculate the result if the *plus* numbers are combined into a single number representing a combination of the upward movements and if the *minus* numbers are similarly combined.

At first the numbers themselves may be rearranged for this purpose. As familiarity with the manipulations is gained the numbers can be combined without rearrangement.

[Ex. XXVII may now be worked.]

B. Algebraic Addition and Subtraction (ch. XVII., §§ 2, 3; Ex. XXVIII).

§ 1. *Two Kinds of "Stair Problems".*—We will begin by working on the board two "stair problems". (i) Starting from the landing I take 5 steps upstairs and then turn round and descend 18 stairs. Where am I after the second movement? *Answer:* Thirteen stairs below the landing. Or, by calculation

$$\begin{aligned} s &= +5 - 18 \\ &= -13 \end{aligned}$$

(ii) Starting from the landing I mean to go to a room opening on the staircase 5 stairs above. By mistake I go up 18 stairs. What must I do to reach the proper destination? *Answer:* Descend 13 stairs. To obtain this result by calculation the working must be arranged as follows:—

$$\begin{aligned} s &= +5 - 18 \\ &= -13 \end{aligned}$$

That is to say, to answer the second question by calculation requires exactly the same working as to answer the first. A person coming into the room and ignorant of our discussion would think that, for some reason, we had worked the same problem twice over. Yet the questions were really entirely distinct. In the first you were told two movements which I made in succession and were asked to calculate where I should be at the end of them; while in the second you were told where I should be after two movements and, being given one of them, were asked to calculate the other.

We shall find it convenient to use here the terms *component* and *resultant* already employed in connexion with vectors. (Ex. XIX, B). If we represented movements up and down stairs by straight lines those lines would, in fact, be vectors, differing from the vectors of Ex. XIX merely in being restricted to one direction along which they point either forwards or backwards. Thus in the first of our problems + 5 and − 18 were components and − 13 was a resultant; in the second + 18 and − 13 were components and + 5 was the resultant.

We conclude that calculations involving directed numbers may, although exactly alike, represent attempts to answer two quite different kinds of questions. Expressed briefly these questions are:—(A) Given two components to find their resultant; and (B) Given the resultant and one component to find the other component.

§ 2. *The Problems distinguished by Symbolism.*—Is it possible to set the calculations down in a way which will show which of the two kinds is under consideration?

Suppose the movements to be confined to a single direction—for example, up a ladder down which it is impossible to return. Then the setting down of the calculations would suffice by its form to show what questions have been asked. Thus the form $s = 30 + 17$ could mean only that a person has taken two successive upward movements of 30 and 17 steps and that the question is how far is he now up the ladder. That is, the presence of the *plus* sign implies a problem of the first kind. Similarly the form $s = 30 - 17$ could mean only that he has set out to climb 30 steps and has actually taken 17, the question being what second movement is necessary to carry him to his destination.

It is natural, then, to inquire whether the signs *plus* and

minus cannot still be used to indicate the character of the problem even when the movements involved may be down as well as up, and require, therefore, other *plus* and *minus* signs as labels to show which way they are taken. There is no difficulty, provided that we are careful to distinguish the latter signs from the former—which can be done by the simple device of enclosing each number with its directive label in brackets. Thus the two problems of § 1 may be stated respectively in the forms

$s = (+ 5) + (- 18)$ and $s = (+ 5) - (+ 18)$

which are no longer ambiguous. The *plus* between the bracketed numbers in the former shows it to be a problem in which we are given two component movements, + 5 and − 18, and are asked to find their resultant. The *minus* between the bracketed numbers in the latter shows that it is a problem in which we are told a resultant movement, + 5, and one of the (actually taken) component movements, + 18, and are asked to find the other component.

§ 3. *Rules of Procedure.*—The only difficulty left is to settle the practical rules by which we are to proceed from the setting down of the problem in symbols to the calculation of the answer. Consider once more the problems of § 1. "Common sense" shows that the answer to both problems must be − 13. To obtain this answer from the numbers 5 and 18 we *must* have + 5 − 18 in each case. Thus *from our knowledge of the correct answer* we see that the calculations must be written :—

$s = (+ 5) + (- 18)$	and $s = (+ 5) - (+ 18)$	(i)
$= + 5 - 18$	$= + 5 - 18$	(ii)
$= - 13$	$= - 13$	(iii)

An analysis of these cases suggests the following rules. The *plus* and *minus* signs between the bracketed numbers in (i) are there *simply to show the nature of the problem.* When we proceed to the actual calculation we drop them. If the sign is *plus* the directed number following it (i.e. the second component) is written down as it stands. If the sign is *minus* the sign of the component is changed. We thus obtain line (ii). The final results (iii) are obtained by combining the directed numbers.

These suggested rules must be tested by applying them to specimens of all possible cases. The following are instances.

(1) A person goes down 7 stairs, then turns and comes up 18. Where is he now? This is a problem of the first kind to which the answer is obviously + 11. The statement must take the form $s = (-7) + (+18)$ since the two components are − 7 and + 18. The rule bids us drop the " + " between the bracketed numbers, for it is there merely to show the nature of the problem. It tells us to write the two components, with their signs, just as they stand in the brackets and to combine the directed numbers. Carrying out these instructions we have

$$\begin{aligned} s &= (-7) + (+18) \\ &= -7 + 18 \\ &= +11 \end{aligned}$$

The answer is correct and the rule is, in this case also, justified. (2) A person intended to go to a room 8 stairs below the landing on which he stood, but, by mistake, went down 20 stairs. What movement must he take to reach his intended destination? Here the problem is of the second kind involving a resultant − 8 and a component − 20. The answer to be expected is + 12. Following the rule we have

$$\begin{aligned} s &= (-8) - (-20) \\ &= -8 + 20 \\ &= +12 \end{aligned}$$

That is, we dispense with the *minus* sign whose sole function is to indicate the type of problem, and at the same time reverse the sign of the following component. The answer is correct and the rule, therefore, justified.

Members of the class will themselves suggest further problems involving different combinations of signs. They will state the answers to be expected and verify that the rules always give them. The result in a case in which (for example) the first number is negative, the second positive and numerically greater than the other may clearly be taken as showing what must happen in *all* such cases. When the various typical cases have been explored we may justly conclude that the rules have universal application.

§ 4. *Algebraic Addition and Subtraction.*—The problem represented by such symbolism as $(+14) + (-9)$ may be called "addition" if we remember that it is something very different from ordinary or arithmetical addition which always produces an increased total. We "add" here only in the

sense in which we may say that we "add" the two vectors AC and CB (fig. 21) to obtain the resultant AB, and + 5 is the "sum" in this case only in the same way that AB is the "sum" of the two vectors AC and CB. In other words "adding" means here combining components, and the "sum" is simply the resultant of the combination. We can distinguish, in fact, three distinct kinds of addition: (1) ordinary arithmetical addition, in which the numbers are non-directed and the total is increased by each term added; (2) "algebraic addition"—the kind now before us—in which the numbers indicate movements or distances backwards or forwards along a line from a certain origin and the "sum" of two components is a third movement or distance along the same line which may be either backwards or forwards and either greater or less than both components; (3) "vector addition," in which we deal with lines representing movements inclined to one another. Here the "sum" is a third line representing a movement which is in a direction different from those of the components and either greater or less than either of them.

These considerations show why the rule of procedure in algebraic addition holds good. The *plus* between the bracketed directed numbers implies that they are to be combined, and the combination, as we saw in Ex. XXVII, is effected by taking account of their signs. Thus to proceed with an algebraic addition we simply drop the *plus* sign that separates the components and deal with the components in accordance with their signs.

Answering to the three forms of addition there are three kinds of subtraction: (1) arithmetical subtraction, which always consists in reducing a non-directed total by taking a number away from it; (2) algebraic subtraction, in which we are given a directed number representing a resultant movement along a line from a certain origin together with another directed number representing a component movement along the same line, and are to find a third directed number which will represent the second component—and may be numerically greater than the resultant; (3) vector subtraction, in which we are given a line AB (fig. 21) representing a resultant movement and a line AC representing a component movement, and are to find the second component CB.

Algebraic and vector subtraction differ from arithmetical subtraction in a further important respect. Every problem

of algebraic or vector subtraction can be turned into a problem of addition involving the same numbers or lengths. Thus suppose we are given the vectors AB and CB (fig. 21) and wish to find the vector AC. This is a problem of vector subtraction. Draw AB, and from B draw, not CB but BC, that is, a vector of the same length and direction as CB but with the arrow-head reversed. The figure now represents a problem of vector addition, but the answer, AC, is exactly the same as the answer to the original problem of vector subtraction. We can explain the result by saying that AB, being equivalent to AC and CB, can be replaced by them. When we "add" BC, the addition cancels CB and leaves us with the other component AC.

Consider similarly the problem of algebraic subtraction, $s = (-8) - (-20)$. The resultant (-8) is equivalent to the known component movement (-20) together with the unknown movement s and may therefore be replaced by them. If we add to the combination of s and (-20)—that is, to (-8)—the component $(+20)$, the effect of the movement (-20) will be cancelled [1] and the unknown component s will be revealed. In other words the *addition* problem $s = (-8) + (+20)$ must yield the same answer as the *subtraction* problem $s = (-8) - (-20)$.

We now see the justification of the rule of procedure in algebraic subtraction. A problem of algebraic subtraction can always be replaced by one of algebraic addition in which the second component is the original component with sign reversed. Since the "problem sign" in front of this component is now *plus*, the number suffers no further change of sign when the *plus* is suppressed and the calculation begins. That is, the second term of the subtraction enters into the calculation with sign reversed.

§ 5. *The Use of Symbols.*—Hitherto such an expression as $a + b$ has always meant "addition" in the ordinary sense of the word; a and b have stood for two numbers of which the second was to be added to the first and so to increase it. Similarly $a - b$ has always meant the subtraction of one number from another with a diminution of the total. But henceforward if we are not told that our symbols stand for non-directed numbers it must be assumed that they stand for

[1] Going up 20 stairs cancels the effect of going down 20.

directed numbers and that the operations required are algebraic addition and subtraction.

If we are given the values of a and b in such expressions we now know how to proceed to calculate the resultant or the unknown component as the case may be. As a rule the symbols will be accompanied by numerical coefficients. Here is an example :—

Given that $c = 2a - 3b$, find the value of c when $a = -3$ and $b = -5$.

$$\begin{aligned} c &= 2a - 3b \\ &= 2(-3) - 3(-5) \\ &= (-6) - (-15) \\ &= -6 + 15 \\ &= +9 \end{aligned}$$

The problem here is one of algebraic subtraction. The resultant is given as $2a$, that is, as equivalent to two movements each to be denoted by a. In a similar way the known component is given as $3b$, that is, as equivalent to three movements each to be denoted by b. In passing from the second to the third line of the working we argue that two movements of -3 steps are equivalent to a single movement of -6 steps and that three movements each of -5 steps are equivalent to a single movement of -15 steps. After this we apply the rule for algebraic subtraction.

[Ex. XXVIII, Nos. 1-31, may now be worked.]

§ 6. *More Complicated Cases.*[1]—Given that

$$c = (3a - 4b) + (2a - 5b)$$

find the value of c when $a = -4$ and $b = +3$.

This is obviously a double problem. In the first place the *plus* sign between the bracketed expressions shows that the problem is one of algebraic addition; c is to be regarded as the resultant of two components represented respectively by $3a - 4b$ and $2a - 5b$. But to determine the value of these components we have to solve two problems of algebraic subtraction. The working will begin with the subtraction within the brackets and will proceed as follows :—

[1] §§ 6 and 7 with Examples C, may be omitted at discretion or taken in revision.

$$
\begin{aligned}
c &= (3a - 4b) + (2a - 5b) \\
&= \{3(-4) - 4(+3)\} + \{2(-4) - 5(+3)\} \\
&= \{(-12) - (+12)\} + \{(-8) - (+15)\} \\
&= (-12 - 12) + (-8 - 15) \\
&= (-24) + (-23) \\
&= -24 - 23 \\
&= -47
\end{aligned}
$$

Given that $c = (3a + 2b) - (5a - 2b)$ find the value of c when $a = -3$ and $b = -4$.

Here the problem is fundamentally one of algebraic subtraction: c is the unknown component of a resultant represented by $(3a + 2b)$, the known component being represented by $(5a - 2b)$. The determination of the value of the resultant $(3a + 2b)$ itself involves a subsidiary problem in algebraic addition; the determination of the value of the component $(5a - 2b)$ a subsidiary problem in algebraic subtraction. In working we begin as before with these subsidiary computations.

$$
\begin{aligned}
c &= (3a + 2b) - (5a - 2b) \\
&= \{3(-3) + 2(-4)\} - \{5(-3) - 2(-4)\} \\
&= \{(-9) + (-8)\} - \{(-15) - (-8)\} \\
&= (-9 - 8) - (-15 + 8) \\
&= -17 + 7 \\
&= -10
\end{aligned}
$$

The meaning of each step in these calculations should be elicited. In doing the exercises the teacher should insist upon having the problems worked in full until the significance of the successive operations is thoroughly mastered. Those who are able to do so may then be allowed to shorten the working by omitting (say) the second and third lines.

§ 7. *Shorter Methods of Working.*—The working can often be greatly shortened by a preliminary treatment of the formula (cf. the use of factorization and fractions). The shortening depends on the fact that the algebraic subtraction of a given component always produces the same effect as the addition of a component numerically the same as the former but of opposite sign. This rule may be expressed in symbols by writing

$$a - b = a + (-b)$$

the symbol $(-b)$ being used to indicate the component that has the same number as b but the opposite sign. Applying the rule to the first example we have:—

$$
\begin{aligned}
c &= (3a - 4b) + (2a - 5b) \quad . \quad . \quad . \quad . \quad \text{(i)} \\
&= \{3a + (-4b)\} + \{2a + (-5b)\} \quad . \quad . \quad \text{(ii)} \\
&= 3a + (-4b) + 2a + (-5b) \quad . \quad . \quad \text{(iii)} \\
&= 5a + (-9b) \quad . \quad . \quad . \quad . \quad . \quad \text{(iv)} \\
&= 5a - 9b \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(v)} \\
&= -47
\end{aligned}
$$

The problem has been turned [line (ii)] into one in which we have first to find the resultant of a pair of components, then the resultant of another pair, and, finally, the resultant of these two resultants regarded as components. It is quite obvious that these operations would have exactly the same result as if we simply found the total resultant of the four components [line (iii)] without first finding their partial resultants in pairs. Line (iv) follows obviously. In line (v) the problem is expressed as one of subtraction so that we may use the original component b instead of its opposite. Thus by these manipulations the long expression with which we started is reduced to the simple form $5a - 9b$ before calculations begin. The same method may be applied in the second example :—

$$
\begin{aligned}
c &= (3a + 2b) - (5a - 2b) \quad . \quad . \quad . \quad \text{(i)} \\
&= (3a + 2b) + \{(-5a) - (-2b)\} \quad . \quad . \quad \text{(ii)} \\
&= (3a + 2b) + \{(-5a) + 2b\} \quad . \quad . \quad \text{(iii)} \\
&= 3a + 2b + (-5a) + 2b \quad . \quad . \quad . \quad \text{(iv)} \\
&= -2a + 4b \quad . \quad . \quad . \quad . \quad . \quad \text{(v)} \\
&= -10
\end{aligned}
$$

The subtraction of $(5a - 2b)$ in line (i) is to be replaced by the addition of the opposite component. It is obvious that if we change the signs of both members of $(5a - 2b)$ the component represented will be the one required; in this way we obtain line (ii).

When the arguments are thoroughly understood it may be pointed out that the expressions could be reduced in two steps as follows :—

$$
\begin{aligned}
c &= (3a - 4b) + (2a - 5b) \quad . \quad . \quad . \quad \text{(i)} \\
&= 3a - 4b + 2a - 5b \quad . \quad . \quad . \quad \text{(ii)} \\
&= 5a - 9b
\end{aligned}
$$

$$
\begin{aligned}
c &= (3a + 2b) - (5a - 2b) \quad . \quad . \quad \text{(i)} \\
&= 3a + 2b - 5a + 2b \quad . \quad . \quad . \quad \text{(ii)} \\
&= -2a + 4b
\end{aligned}
$$

The sign in front of a bracket is to be prefixed to the first number within the bracket. The sign in the bracket remains unchanged if the sign before the bracket is *plus* and is changed if that sign is *minus*.

The rules exemplified by the transition from line (i) to line (ii) are called the "law of association". Compound quantities may be resolved into their elements by the removal of brackets—accompanied by a change of sign whenever there is a *minus* before the bracketed group of numbers.

[Ex. XXVIII may now be finished.]

C. The Multiplication and Division of Directed Numbers (ch. XVII., §§ 2, 4; Ex. XXIX).

§ 1. *Train Problems; Graphic Solution.*—Suppose that we were standing on the platform of the railway station at Doncaster at the moment when a G.N.R. express passes through on its way from London to the north. Let us suppose, also, that the train keeps up a uniform speed of 42 miles an hour. Then it would be easy to draw a graph which would represent its position at any moment after it passed us. (The graph is the line OP of fig. 47, the axes being OT and OD.)

Now this train had a history before it reached Doncaster, and we might, of course, draw a second graph to represent it. But it would clearly be better if we could extend the present graph so as to include this history; that is, so as to show the position of the train at *any* moment of its journey, either before or after it reached Doncaster. This can be done very simply. Produce the axis TO to T′, and the axis DO to D′. Graduate OT′ from O to represent times before the train reached Doncaster, and graduate OD′, also from O, to represent distances from Doncaster towards London. Then if we calculate the positions of the train 1, 2, 3 . . . hours before it reached Doncaster and insert points in the extended graphic scheme to represent them, these points will mark out a straight line, OQ, which is obviously a continuation of OP.

We have now a graph showing the position of the train at all moments of its journey. To make it completely serviceable it will be well to indicate in some way the facts that horizontal distances to the right measure times *after* and distances to the left measure times *before* the train reached

Doncaster ; also that distances above and below the line TT′ measure respectively distances of the train to the north or the south of Doncaster. The simplest way to do this will be to

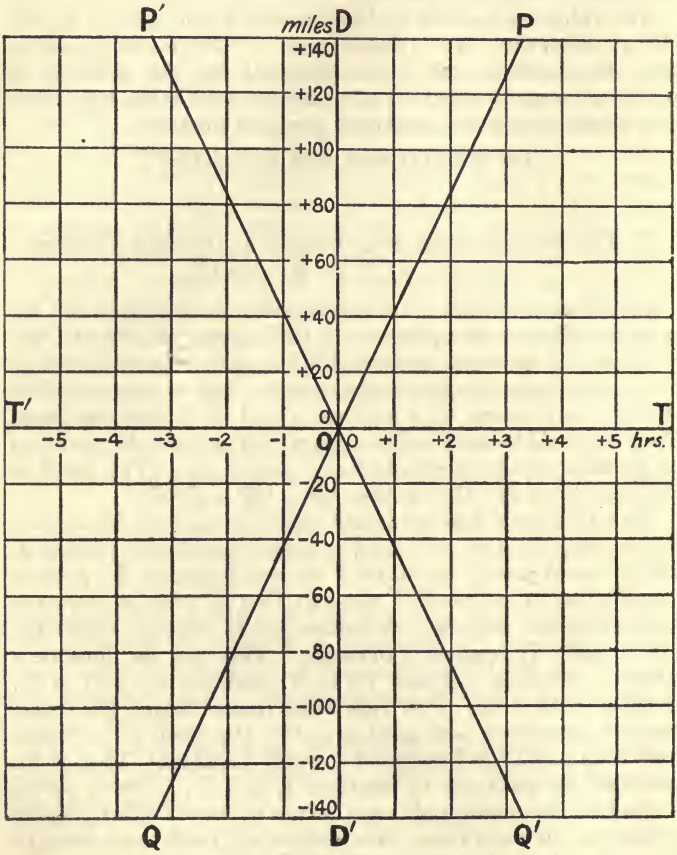


Fig. 47.

distinguish these times and distances by directed numbers. It does not matter at all which are reckoned positive and which negative, but it is usual, in drawing a graph, to mark distances to the right or upwards, positive, to the left or

downwards, negative. Thus we are going to count as positive the times after the train reached Doncaster and its distances from Doncaster after it has passed through the station. The other times and distances will be reckoned negative.

§ 2. *Graph of a Southward Journey.*—Next suppose that at the moment when this train is passing through Doncaster on its way north another train is also passing through and at the same speed of 42 miles an hour, but towards London. How will the graph representing its positions at various times differ from the one just constructed? To answer this question we will plot a few of the points that it must contain. Thus, 2 hours before it reached Doncaster it would be 84 miles to the north; the point representing this fact must be placed on a level with the graduation + 84 on the axis DD′ and directly above the graduation − 2 on the axis TT′. Again, 3 hours after passing us the train would be 126 miles to the south. The representative point will this time be on a level with the graduation − 126 in the distance scale and below the graduation + 3 on the time scale. One or two more points having been fixed in the same way, the graph comes out as a straight line, P′Q′, making with the axes DD′ and TT′ exactly the same angle as the former line but lying on opposite sides of both of them.

From the two graphs thus drawn with the same axes and the same scales it is possible to read off, without calculation, the answers to a number of problems: e.g. How far apart are the two trains $3\frac{1}{4}$ hours before they reach Doncaster? What length of time elapses between their passages through York (+ 32 miles)?

§ 3. *The Position of the Train by Calculation.*—We should have, of course, no difficulty in obtaining answers to any of these same problems by calculation. The familiar relation

distance covered by train = speed × time taken

would enable us to deal easily with all of them. But in this formula, as you have hitherto used it, the distance, time and speed have not been thought of as directed numbers. The question may now be considered whether it is possible to bring directed numbers under its operation. That is to say, can the formula be made to contain and impart all the information about the direction (as well as the magnitude) of the speed, time and distance which is given us by the lines upon the graph paper?

In the first place we will mark the fact that we wish our calculation to show which way the train is going as well as how fast it is moving by replacing the word "speed," which refers only to the train's rate of movement, by the word "velocity," which is generally taken to refer also to the direction of movement. When the train is moving northwards the velocity will be reckoned positive, when southwards, negative. Our formula may now be written in the form $d = vt$, and it is to be understood that all three symbols stand for directed numbers.

Let us now attempt to work out some examples by means of the formula.

(*a*) Where will the northward moving train be 3 hours after it leaves Doncaster? Here $v = +42$, $t = +3$, and *we see from the graph* that $d = +126$. Thus our calculation must take the form :—

$$\begin{aligned} d &= vt \\ &= (+42) \times (+3) \\ &= +126 \end{aligned}$$

That is, the multiplication of the two positive directed numbers must be supposed to yield another positive number. It is clear from the graph that this rule will hold for all cases in which a positive speed is to be multiplied by a positive time; for the products are all represented by points on the line OP —that is, by points which represent positive distances from Doncaster.

(*b*) Where will the same train be 3 hours before it reaches Doncaster? Here $v = +42$, $t = -3$, and *we see from the graph* that the distance is -126. Hence we must have :—

$$\begin{aligned} d &= vt \\ &= (+42) \times (-3) \\ &= -126 \end{aligned}$$

That is, the multiplication of the positive by the negative directed number yields a negative number. Again we see that the rule always holds good. For the points which represent all such products lie on the line OQ which represents places whose distances from Doncaster are negative.

(*c*) Where will the southward-moving train be 3 hours after it reaches Doncaster? Here $v = -42$, $t = +3$, and we see from the graph that the distance is -126. Thus we must have :—

$$d = vt$$
$$= (-42) \times (+3)$$
$$= -126$$

That is, the multiplication of the negative number by the positive number yields a negative number. Since all such products are on the line OQ′, this rule always holds good. It shows, too, that directed, like ordinary numbers, follow the "law of commutation" in multiplication—that is, that the multiplier may become the multiplicand and conversely without change in the value of the product. We could not have been sure, because $3 \times 4 = 4 \times 3$, that $(-3) \times (+4) = (+4) \times (-3)$; but we have found that it is so.

(*d*) Lastly, Where will the southward-moving train be 3 hours before it reaches Doncaster? Here $v = -42$, $t = -3$, and we see from the graph that $d = +126$. Thus we must have:—

$$d = vt$$
$$= (-42) \times (-3)$$
$$= +126$$

That is, the multiplication of the negative number by another negative number yields a positive number. As in the other cases the fact that the points representing such products all lie upon OP′ shows that the rule is universally true.

§ 4. *Comparison with former Results; the "Rule of Signs".*—The rules just found for replacing two signs by a single sign when two directed numbers are multiplied are exactly the same as those which we discovered when we were working with the formulæ $c = a + b$ and $c = a - b$. We found that such combinations as $+(+4)$, $+(-4)$, $-(+4)$ and $-(-4)$ can be replaced by the single directed numbers $+4$, -4, -4, $+4$, respectively; the rule being that the *plus* sign of the formula may be dropped without affecting the sign of the directed number, while the *minus* sign of the formula could be dropped only if at the same time the sign of the directed number was reversed. These rules, those of the present lesson, and the rules of arithmetical addition and subtraction of ch. VIII., B, can all be summed up in one: "Like signs produce *plus*; unlike signs produce *minus*". This is called "the rule of signs".

It is true that we have arrived at these rules from the consideration of only one type of example in each case; but there can be no doubt that they will hold good whenever we

are dealing with magnitudes that can be represented graphically to the right and left or above and below a zero point. We should not doubt that three sevens always made two tens and a one simply because we had discovered the truth only in counting up piles of pebbles. It is obvious that the result must be true of any things that can be gathered into groups, even though we have never tried it with them. For the same reason we may be sure that the rule of signs can be trusted in all cases in which a graph like that of the present lesson can be used to represent the magnitudes in question.

§ 5. *Division of Directed Numbers.*—It can easily be shown that the rule of signs will hold good for division of directed numbers as well as for multiplication. When we say that $\frac{12}{3} = 4$ we mean, of course, that 4 is the number by which 3 must be multiplied in order to produce 12. In the same way, if I inquire the value of $\frac{-12}{-3}$ I mean to ask what -3 must be multiplied by in order to yield the product -12. In accordance with the rule of signs the answer is $+4$; that is, in division two *minus* signs will be replaced in the quotient by a *plus*. The other cases of division can be treated in the same way.

[Ex. XXIX may now be worked.]

CHAPTER XIX.

CONSTANT-DIFFERENCE SERIES.

A. The Summation of Constant-Difference Series (ch. XVII., § 5; Ex. XXX, A).

§ 1. *The Problem.*—Among the "events" of school athletic sports the "Block Race" frequently finds a place. Each competitor, as he stands on the starting line at O, has in front of him a number of wooden blocks, placed at equal distances at the points A, B, C, etc. When the signal is given he has to run from O to A and return with the block to O. Leaving it there he runs to B, fetches the second block and, returning again to O, places it upon the first. He then runs to C, returns to O with the third block, and places it on the top of the second. The race continues in this way until one competitor has fetched in and piled up all his blocks. Suppose that the distance OA is 4 yards and that the blocks are 3 yards apart, then the lengths of the various journeys of a competitor, measured in yards, form the sequence:—

$$8, 14, 20, 26, 32, \ldots$$

This is evidently an arithmetical progression with first term 8 and constant difference 6. To find the total distance run in the race we could, of course, simply add the successive terms of the series together. But if it was a very long race this addition would be a tedious business. We will try, therefore, to find an easier way of arriving at the answer to our problem.

§ 2. *Summation Formula; (i) when the Terms are Non-directed Numbers.*—Fig. 48 is a column-graph representing the distances covered in the first seven journeys. The area of the shortest column represents 8 yards, that of the longest

$$8 + 6 \times 6 = 44 \text{ yards.}$$

The total distance run is represented by the total area of the figure. Since we have chosen an odd number of strips

there must be one, marked P in the figure, which has an equal number of others to right and left of it. Moreover, since the strips increase uniformly in height, it is evident that if the piece marked q were cut off from the strip Q it would fit exactly into the space q' on the top of Q'. Similarly r and s would fit into the spaces r' and s'. In this way the figure could be turned into a rectangle, consisting of seven strips of the same height (fig. 49). If we inquire what that height is, we note that the two strips S and S' of fig. 48 together make *two* of the equal strips of fig. 49. Each of these strips is, therefore, one half of the sum of the first and last strips of fig. 48—that is $\frac{1}{2}$ of (8 + 44), or 26. Since there are seven such strips, the total area of the figure—the total distance run—is 26 × 7, or 182 yards. Even with so small a number as seven journeys this way of performing the calculation is quicker than addition. But it is obvious that the argument could be applied equally well to *any* odd number of journeys. For every piece cut off from a strip to the left of P there will be an equal space to be filled above a strip on the right so as to convert the whole area into a rectangle composed of equal strips. Moreover, each of these equal strips must always be one half of the sum of the first and last of the original strips. For example, suppose that the race consisted in fifty-one journeys, so that the last would be one of 8 + 50 × 6 = 308 yards. Then the total distance would be represented by a rectangle composed of fifty-one equal strips each of area $\frac{1}{2}(8 + 308)$, i.e. 158. Thus the total length of the race would reach the formidable total of 158 × 51, or 8058 yards, that is about $4\frac{1}{2}$ miles!

We conclude that when n is an odd number, a the first and l the last of the terms, the sum S is given by the formula

$$S = \frac{n}{2}(a + l).$$

Consider next the case of an even number of terms of the series, say 8 (fig. 50). This time there is no middle strip, so the 5th and 4th strips are marked P and P'. But if we draw a line parallel to the base half-way between the top of P and and the top of P' it is clear that the piece p will fit into the space p', the piece q into the space q', etc., as before. The figure will once more be converted into a rectangle made up of equal strips and once more each of these strips will be one

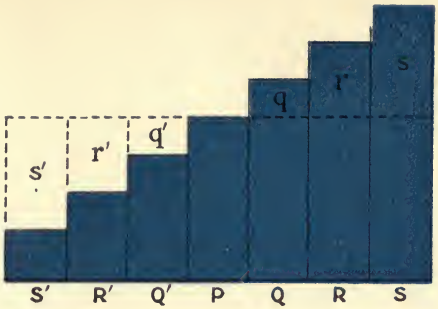


FIG. 48.

FIG. 49.

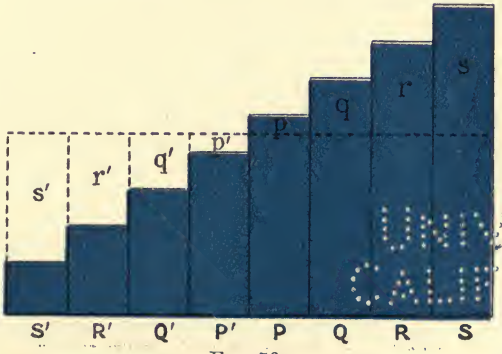
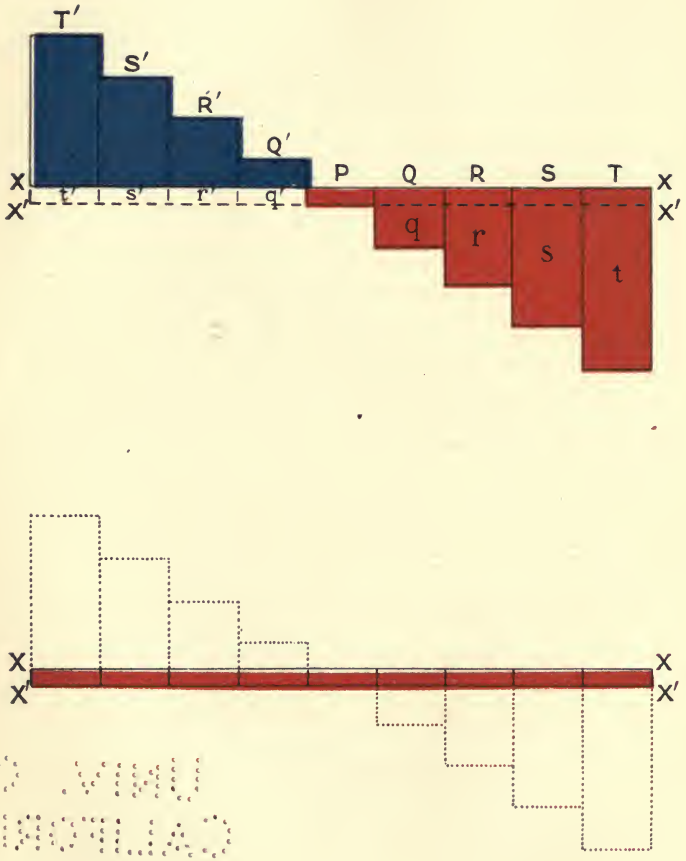


FIG. 50.

FIG. 51.



half the sum of the first and last of the original strips. Thus the area of the rectangle, and consequently the total distance run in the race, will in this case also be given by the formula

$$S = \frac{n}{2}(a + l).$$

§ 3. *The Summation Formula; (ii) when the Terms are Directed Numbers.*—It is natural to inquire whether this formula holds good for constant-difference series composed of directed numbers, some of them negative. Let us examine the case of the series

$$11 + 8 + 5 + 2 - 1 - 4 - 7 - 10 - 13 \ldots\ldots$$

As before we will begin with an odd number of terms—say 9—and will represent their values by the areas of strips of constant width.

The first four (positive) terms will be represented as before by the decreasing strips T′, S′, R′, Q′, of fig. 51. We then come to the negative terms. It is quite natural to represent these by strips P, Q, R, S, T, drawn below the base line XX. The principle we are adopting is that if the area of a rectangle is taken as representing a positive directed number the area of the same rectangle when inverted can be taken to represent the corresponding negative number. It will be noticed that upon this plan the horizontal ends of the strips descend in height by uniform steps along the diagram. Since the bottom line of a negative column corresponds to the top line of a positive column the column must be supposed to become negative not by slipping down the face of the paper but by revolving out of the plane of the paper about XX (fig. 53). If it is coloured blue on one side and red on the other, the change of sign will be signalized by a change of colour.

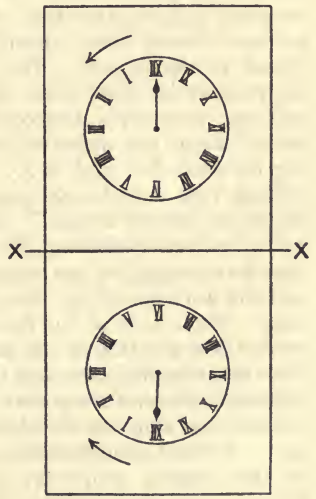


FIG. 53.

Suppose the strip to contain a

transparent clock-dial, and that when it is in the positive position we are standing behind the dial. Then when the strip revolves about XX the dial will be facing us (though upside down) and the hands will be going round in the ordinary way. This idea may be used to distinguish between areas that are to be measured by positive and negative numbers when we cannot conveniently distinguish them by colours. An area measured by a positive number may be distinguished by a curved arrow (fig. 53) indicating movement in the counter-clockwise direction; the area measured by a negative number being marked by an arrow indicating clockwise movements.

Since P in fig. 51 is the middle strip we will, as in the former case, draw the dotted line X′X′ at the level of its end. In this way we cut off from each of the strips Q, R, S, T, a piece equal to P. Moreover, it is clear that the residue from Q, marked q, would, if laid upon Q′ cover the whole of it and project below XX to the line X′X′. Similar statements are true with regard to the pieces of R, S, and T marked r, s, t. But q, r, s, and t are negative areas. If they are laid upon positive areas we must suppose that they obliterate or cancel any positive surface beneath them, disappearing themselves to an equal extent in the process. Thus q, r, s and t, entirely wipe out the positive (or blue) areas Q′, R′, S′, T′, above the base XX and are themselves destroyed to the same extent, but leave behind them the negative (red) spaces q', r', s', t', outstanding between XX and X′X′. As before the figure is transformed into a rectangle (fig. 52) made up of equal strips. Moreover, out of T and T′, the first and last of the original strips, we have remaining, after the equal positive and negative ones have cancelled one another, two of these equal strips, t' and the corresponding piece above t at the other end of the base. Thus each of the equal strips is one half of the sum of the first and last of the original terms. The only difference between this case and the former cases is that the areas represent directed numbers and that the "sum" is not their arithmetical sum but the algebraic resultant of their combination. In the instance before us the areas of the first and last of the original strips are + 11 and − 13 respectively. It follows that the common area of the nine equal strips constituting the rectangle XXX′X′ is $\frac{1}{2}\{(+ 11) + (- 13)\}$, i.e. − 1. The total resultant area and the sum of the series

will, therefore, be $(-1) \times 9 = -9$. This result can be verified by addition. We conclude that in this case also the sum of the series is given by the formula

$$S = \frac{n}{2}(a + l)$$

a and l now being directed numbers.

§ 4. *Proof by Symbols.*—We could examine all other possible cases in the same way but there can be no doubt that the formula is universally valid. The essence of the argument is to show that in each case we get a number of equal rectangles—just as many rectangles as there are strips or terms, that the first and last of the original strips together make two of the equal rectangles, that the second and last but one make two more, and so on. We can represent this argument in symbols in such a way as to prove that it always holds good. In one line we write the series forwards and in a second line we write it backwards, so that the last term comes under the first, the last but one under the second, etc.

$$\begin{aligned} S &= a + (a + d) + (a + 2d) + (a + 3d) + \ldots \\ &\quad + (a + n - 1d) \\ S &= l + (l - d) + (l - 2d) + (l - 3d) + \ldots \\ &\quad + (l - n - 1d) \\ 2S &= (a + l) + (a + l) + (a + l) + (a + l) + \ldots \\ &\quad + (a + l) \\ &= n(a + l) \end{aligned}$$

$$\therefore S = \frac{n}{2}(a + l).$$

On addition we obtain an $(a + l)$ for each term of the series —a result which corresponds, of course, to the n equal rectangles of our graphic method. The result $S = \frac{n}{2}(a + l)$ follows at once.

The advantage of the proof by symbols is that a, l and d may be taken to represent directed as well as ordinary numbers and the addition may be algebraic as well as ordinary addition. Hence the one argument by symbols covers all the cases that had to be treated separately by the graphic method.

B. The Calculation of certain Areas and Volumes (ch. XVII., § 6; Ex. XXX, B, C).

§ 1. Fig. 54 represents a model consisting of six white cardboard rectangles all of the same size, and of a group of grey

rectangles resting on them. The grey rectangles increase in size uniformly from b up to c which is as large as the white rectangle below it. To the vulgar eye there are only five grey rectangles but the mathematical eye will discern upon a a sixth whose size is zero! What ratio does the total area of the grey rectangles bear to that of the large white rectangle built up of the equal strips below?

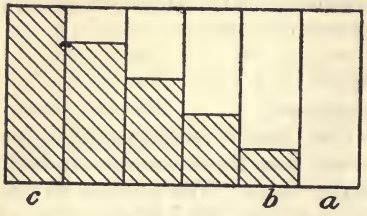


FIG. 54.

We can take the area of grey rectangle b as unity. The area of the largest grey rectangle will then be 5, and this will also be the area of each of the equal white rectangles. The area of the imaginary grey rectangle (a) is, as we have already said, zero. The required ratio is given by the fraction

$$\frac{0+1+2+3+4+5}{5+5+5+5+5+5} = \frac{15}{30} = \frac{1}{2}$$

When we turn to fig. 55 we see a model just like that of fig. 54, with the difference that, although the white strips cover the same total area they are thinner and more numerous. There are 12 of them.

It will be convenient again to take the smallest of the visible grey rectangles as the unit of area. (The change of unit does no harm since we are seeking only a ratio of areas.) The largest grey, and each of the equal white, rectangles will, therefore, have an area of 11. The ratio of the grey area to the white will be given by the fraction

$$\frac{0+1+2+\ldots+11}{11+11+11+\ldots+11} = \frac{66}{132} = \frac{1}{2}$$

We need not trouble to set down all the figures of the numerator. They form an A.P. of 12 terms and the sum of the first and last is 11. Thus the sum of the series is

$\frac{11}{2} \times 12 = 66$. The denominator is $11 \times 12 = 132$. Thus we again find that the ratio is one half.

It is easy to show that, however numerous the rectangles, the grey area will always be one half of the white area below. Suppose the total white rectangle to remain of the same size but to be built up of $(p + 1)$ equal strips where p is any number you please. There will be p visible grey strips and in addition the invisible one of zero area. Calling the area of the smallest visible grey strip 1, the area of the largest will be p. This will also be the area of each of the white strips. Thus for our ratio we have

$$\frac{0 + 1 + 2 + \ldots + p}{p + p + p + \ldots + p} = \frac{\frac{1}{2}(p + 1)p}{(p + 1)p}$$

$$= \frac{1}{2}$$

§ 2. *Application of Result to Area of Triangle.*—It appears, then, that no matter how numerous the rectangular strips the area covered by the grey ones is one half of the whole area covered by the white ones. But if the strips, while still covering the same area, became immensely numerous and correspondingly thin, a time would come when it would be impossible to distinguish their corners either by eye or touch. To sight and feeling they would be indistinguishable from this cardboard triangle. But since the ratio of areas has remained one half all along, we conclude that the area of the triangle must itself be one half of the area of the rectangle—or so little different from one half that no one could ever estimate the degree of difference.

§ 3. *Other Applications.*—Of course we already knew that the area of a triangle is one half of that of the corresponding rectangle. But this mode of discovering it is particularly useful because it can be applied to the measurement of many other areas. For example, by moving the grey rectangles of fig. 55 they can be made to present fig. 56. By being moved they are not, of course, changed in area. The total grey area remains, therefore, one half of the big white rectangle and this will be true however numerous the grey rectangles may become. But if I make them numerous enough they will form an area indistinguishable from *Exercises*, fig. 33. There are no rectangles visible in this case, but if I measure the horizontal

distance across the figure at any equidistant intervals, the heights obtained will be in A.P. I shall be measuring, as it were, the heights of submerged rectangles picked out at equal intervals along the line. Whatever intervals I choose, these heights must form an A.P. if the whole of the series is in A.P. (Ex. XXIX, No. 38). The conclusion follows that the area of *Exercises*, fig. 33, is one half of the rectangle whose length is the perpendicular distance from its apex to the bottom boundary and whose width is equal to the length of that boundary.

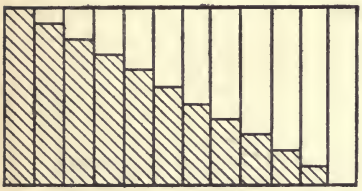


FIG. 55.

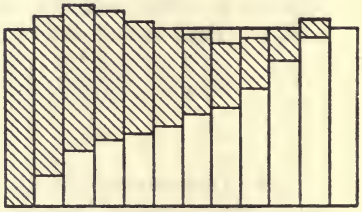


FIG. 56.

Exercises, fig. 34, represents a shape cut out in paper and containing a hole. If any number of equidistant lines are drawn across the figure their lengths (missing out the parts that bridge the hole) will be found to be in A.P. It follows that the shape may be thought of as produced by placing an immense number of rectangular strips in A.P. side by side in certain positions. (Some of the strips must, of course, have been cut and their segments separated so as to leave the hole.) We conclude from this discovery that the area of the shape may be calculated by the same process as that of fig. 33.

[Exercise XXX may now be worked.]

CHAPTER XX.

ALGEBRAIC MULTIPLICATION.

A. Algebraic Multiplication (ch. XVII., § 7; Ex. XXXI, A, B, C).

§ 1. *Are the Identities of Chs. VIII and IX true of Directed Numbers?*—In ch. VII. we saw that formulæ could often be simplified for the purpose of calculation by means of the identities

$$ab \pm ac = (a \pm b)c \text{ and } a^2 - b^2 = (a + b)(a - b).$$

As there used the symbols a, b, c referred to non-directed numbers, and the signs *plus* and *minus* indicated ordinary arithmetical addition or subtraction. That being the case it was easy to show that the identities hold good for non-directed numbers whatever measurements are intended by them. But if a, b, c are symbols for directed numbers the *plus* and *minus* signs in $(a + b)$ and $(a - b)$ imply *algebraic* addition and subtraction, and the former proof of the identities is no longer sufficient. To determine whether they may be used to simplify formulæ expressed in directed numbers requires a fresh investigation.

§ 2. *The Method of Proof; Use of Directed Areas.*—The identities could be tested by substituting in them all kinds of directed numbers, positive or negative, large or small, and seeing if they worked out correctly or not. But this is not a very satisfactory kind of proof, for it would not show *why* the results are correct or incorrect. It will be better, therefore, to seek magnitudes which can be represented by the identity which we are testing and see whether the identity expresses what we know to be the behaviour of the magnitudes represented. It was in this way that we found out the rule of signs in multiplying directed numbers. We used the directed numbers to describe the behaviour of a moving train—a thing with

which we are quite familiar—and so found out what must be the laws of their combination when multiplied together.

We first discovered the identities in question in connexion with the calculation of areas. The use of *directed* areas in the last chapter suggests that the old way of investigation, suitably modified, may again give what we require.

When we thought of the product $(a + b)c$ or the product $(a + b)(a - b)$ as measuring areas, the two factors were regarded as measuring the lengths of the adjacent sides of a rectangle. We can retain this idea with the difference that $a + b$ (for example) must be a *directed* length, the algebraic sum of two directed lengths or measurements, a and b. Thus if $a = -7$ and $b = +3$ then the side of the rectangle measured by $a + b$ must be -4, and our diagram must represent this fact. Similarly if $c = +3$, then the other side of the rectangle must be marked so as to indicate the direction as well as the magnitude of this number. We will adopt the rule always to draw the two directed sides *from* one of the corners of the rectangle. With this condition the product $(a + b)c$ or $(-4) \times (+3)$ would be represented by one or both of the following figures :—

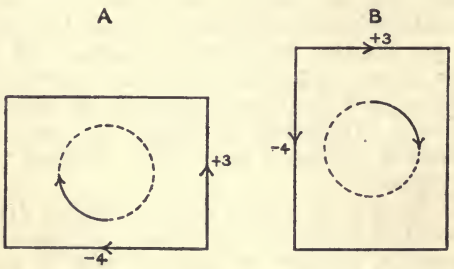
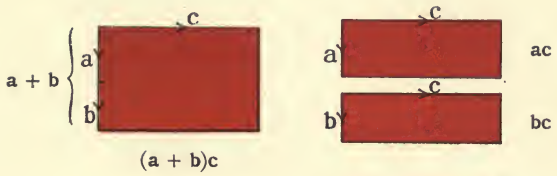


FIG. 57.

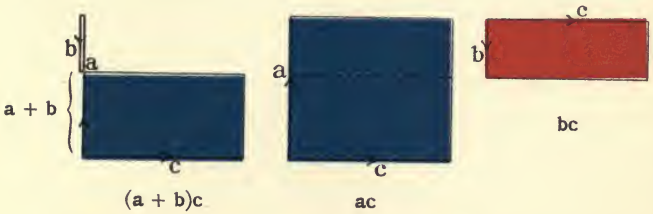
Now if these figures differ at all in what they represent, one of them must stand for $(+3) \times (-4)$, the other for $(-4) \times (+3)$. But, as we have seen, these two products are both -12. We must, therefore, fix the sign of the areas A and B by a rule which makes them both negative. The following rule will be found to produce the required result. Imagine the area to contain a clock-face (see p. 201) with the

FIG. 58.

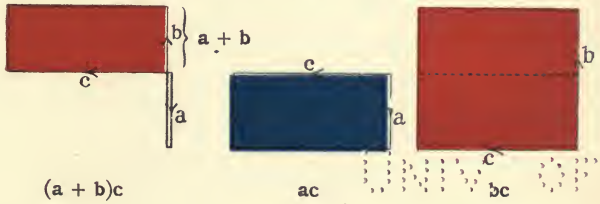
$(a + b)c = ac + bc.$
a and b both negative, c positive.



a positive, b negative, c positive ; a numerically greater than b.



a negative, b positive, c negative ; a numerically less than b.

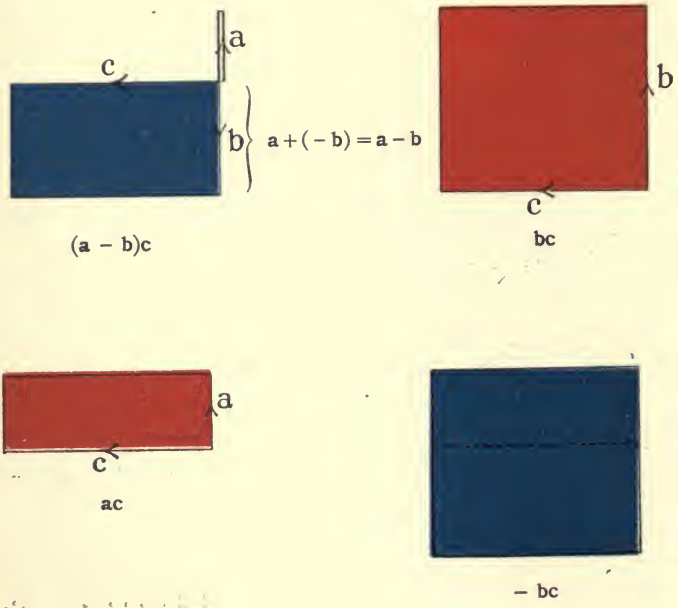


UNIV. OF CALIFORNIA

FIG. 59.

$(a - b)c = ac - bc.$

a and b both positive, c negative; a numerically less than b.



hand pointing to the *marked* horizontal side. Place your finger on the tip of the imaginary hand and move it in the direction indicated by the arrow on the horizontal side. As your finger approaches the vertical side it will turn upwards or downwards according to the direction of the arrow on the other vertical side. It will be seen that in both cases (as indicated by the curved arrow) the hand has been turned the same way and that that way is the "clock-wise" direction which we agreed in ch. xv., A, to call negative. Thus each of the figures A and B will equally well represent either of the products $(+ 3) \times (- 4)$ or $(- 4) \times (+ 3)$. It will, however, add to clearness to restrict each figure to the representation of one product. We will adopt the rule that the base shall represent the multiplier and the vertical the multiplicand. Thus fig. A will represent $(+ 3) \times (- 4)$.

The class should suggest and verify the rules of representation in other cases.

§ 3. *Examination of* $(a \pm b)c = ab \pm ac$.—This idea may now be applied to the examination of identities, beginning with $(a + b)c$. Various cases are possible; three of them are worked out in the diagrams of fig. 58. As in the last lesson, positive areas are coloured blue, negative areas red. When a blue area is superimposed upon a red area or a red area upon a blue one they destroy one another to the extent of their coincidence. In each set of diagrams there is first represented the area $(a + b)c$, $a + b$ being represented by the vertical and c, the multiplier, by the horizontal side. Next we have the areas ac and bc separately. It is obvious that when these areas are brought together their algebraic sum is always identical, both in extent and in sign, with the area $(a + b)c$.

In testing $(a - b)c$ (fig. 59) we have to face the difficulty of representing the algebraic difference $a - b$ at the beginning of the investigation, and $ab - ac$ at the end. The difficulty is overcome by remembering that the algebraic subtraction of a component b is identical in result with the addition of a component $- b$ (p. 189). Thus we start by drawing the rectangle one of whose sides is $a + (- b)$, and when we have obtained the area bc we take the area $- bc$, equal to the former but opposite in sign, and proceed to add it algebraically to the area ac. It will be seen that in each case the identity is verified. Thus in fig. 59 the red ac will cancel

the upper part of the blue $- bc$, leaving a blue area identical with the original rectangle.

The other possible cases could all be tested in the same way, and it then becomes certain that we may in all circumstances assert the truth of the two identities $(a \pm b)c = ac \pm bc$.

§ 4. *Examination of* $(a + b)(a - b) = a^2 - b^2$.—In figs. 60, 61 the identity $(a + b)(a - b) = a^2 - b^2$ is tested in a similar way. We begin by drawing an area whose sides represent the factors $a + b$ and $a - b$ respectively. We then draw the area a^2 and the area b^2. (Note that they must be positive, whether a and b are themselves positive or negative.) The latter area with sign reversed represents $- b^2$. The algebraic addition of $- b^2$ to a^2 must in each case give a figure like fig. 16, for a^2 is always positive and $- b^2$ always negative. The only difference possible in different cases is the "colour" of the residual area. The diagrams show that it has always the same colour, i.e. the same sign, as the area $(a + b)(a - b)$ while by the method of dissection familiar to us from ch. VII., B, it can be shown to have the same extent. Thus in every case $(a + b)(a - b) = a^2 - b^2$.

§ 5. *Identities Proved by Multiplication.*—In ch. VII. we were not content until we had proved by multiplication that the identities are universally true. It is important to find whether the multiplication-process can be applied to combinations of symbols in which the letters stand for directed numbers and the connecting signs imply algebraic addition or subtraction. Fig. 62 illustrates the analysis by multiplication of the identity $(a + b)(a - b) = a^2 - b^2$. The first line represents the argument

$$(a + b)(a - b) = (a + b)\{a + (- b)\} \quad . \quad . \quad \text{(i)}$$
$$= (a + b)a + (a + b)(- b) \quad . \quad . \quad \text{(ii)}$$

Below, the two terms of (ii) are analysed respectively into $(a + b)a = a^2 + ba$ and $(a + b)(- b) = - ab + (- b^2)$. (iii) The elements within the dotted rectangle are now "collected". The terms ba and $- ab$ cancel, and $a^2 + (- b^2)$ yields the familiar figure representing $a^2 - b^2$. The three horizontal lines of figures answer, therefore, to the first three steps of the process

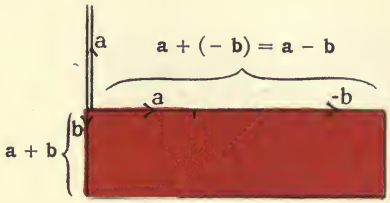
$$(a + b)(a - b) = a^2 + ba$$
$$- ab - b^2$$
$$= a^2 - b^2$$

First there is the (mental) analysis of the multiplier $a - b$

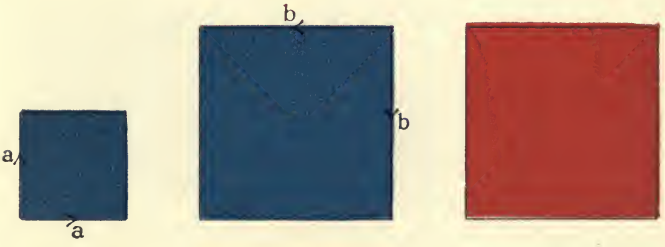
FIG. 60.

$(a + b)(a - b) = a^2 - b^2$.

a positive, b negative ; a numerically less than b.



$(a + b)(a - b)$



a^2 b^2 $- b^2$

$a^2 + (- b^2) = a^2 - b^2$

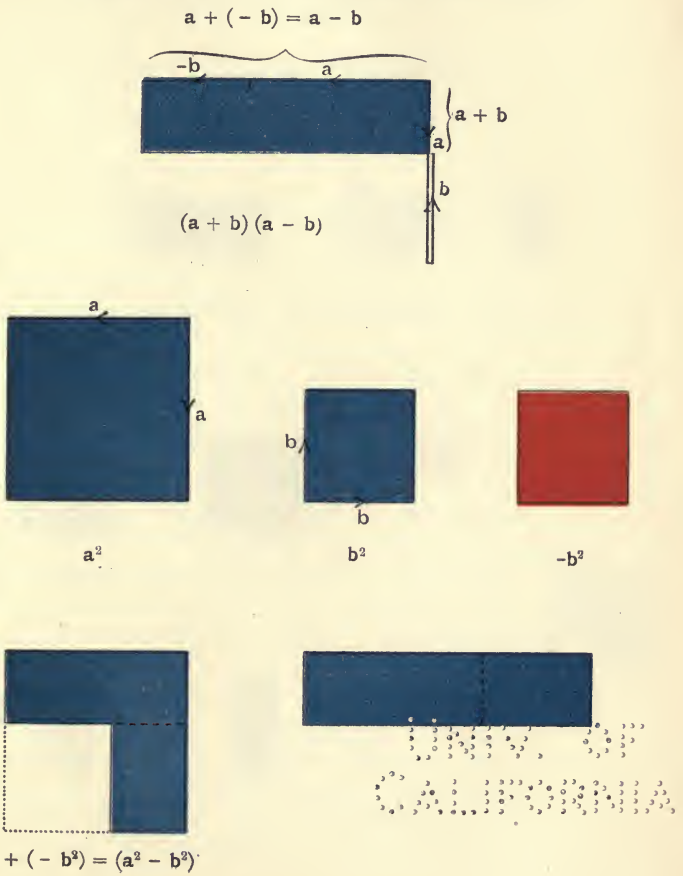
UNIV. OF CALIFORNIA

UNIV. OF
CALIFORNIA

FIG. 61.

$(a + b)(a - b) = a^2 - b^2$.

a negative, b positive; a numerically greater than b.



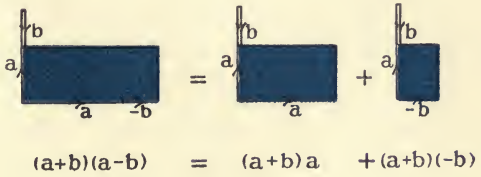
UNIV. OF
CALIFORNIA

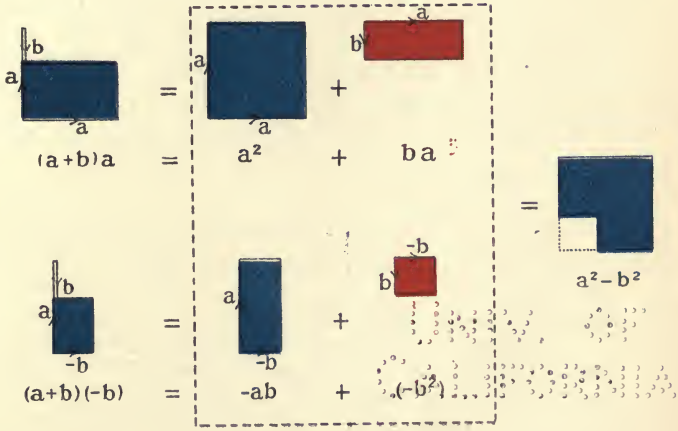
FIG. 62.

$(a + b)(a - b) = a^2 - b^2$.

a positive, b negative ; a numerically greater than b.

$$(a + b)(a - b) = a^2 + ba - ab - b^2 = a^2 - b^2.$$





UNIV. OF
CALIFORNIA

into two multipliers a and $-b$; and then the successive expansion of the products $(a + b)a$ and $(a + b)(-b)$. Finally in each case comes the collection of terms.

The frontispiece of *Exercises*, I, gives an analysis of a special case of $(a + b)(a + b)$ exhibiting exactly similar stages.

A study of these cases leaves no doubt that the multiplication process can be applied universally. The multiplier can always be expressed as an algebraic sum by the device illustrated in line (i). The analysis into a series of partial products, of which two cases are represented by the top lines of fig. 62 and *Exercises*, frontispiece, can then always be carried out. If there are n terms gathered together in the multiplier there will be n of these partial products. Analyses, such as those of figs. 58 and 59, show that the partial products can then always be expanded into their elements by applying the ordinary rule of signs. There remains the collection of terms which is, of course, nothing more than a counting up of the elementary terms of each sort which have been produced by the expansion.

The whole process may be called "algebraic multiplication". This term is meant to remind us that the factors are *algebraic* sums and differences, not arithmetical. The discovery that algebraic multiplications can be carried out as if they were multiplications of arithmetical sums and differences is a proof that the signs *plus* and *minus* when they indicate the algebraic addition and subtraction of directed numbers are subject to the familiar rule of signs.

[Ex. XXXI, A, B, C, can now be worked.]

B. The Binomial Expansion (ch. XVII., § 7; Ex. XXXI, D).

§ 1. *Stifel's Table.*—The results obtained by expanding the products $(a + b)(a + b)$, $(a + b)(a + b)(a + b)$ or $(a + b)^3$, $(a + b)^4$, $(a + b)^5$, etc., have special interest because each can be derived from its predecessor in a very striking way.

There are two things to determine in these expansions or distributions: first the succession of literal terms and secondly the numerical coefficients that go with them. As an example consider the derivation of the expansion of $(a + b)^4$ from that of $(a + b)^3$

$$(a + b)^4 = (a^3 + 3a^2b + 3ab^2 + b^3)(a + b).$$

The multiplication by the term a will produce a series of

terms in which the literal elements are a^4, a^3b, a^2b^2, ab^3—that is, the terms of $(a + b)^3$ with the a raised to the next higher power in each. Multiplication by b produces the series a^3b, a^2b^2, ab^3, b^4—that is the same series again, with the exception that the first term a^4 is missing while an end term b^4 is added. On the whole, therefore, we shall have a gamut of terms from a^4 to b^4 running through intermediate terms in which the power of a constantly falls by unit steps and the power of b constantly rises by unit steps. It is obvious that if we proceeded from the expansion of $(a + b)^4$ to that of $(a + b)^5$, from that of $(a + b)^5$ to that of $(a + b)^6$ and so on indefinitely, this feature would constantly be reproduced. Thus in the expansion of $(a + b)^8$ we must expect the terms a^8, a^7b, a^6b^2, . . . ab^7, b^8.

Having considered the literal terms apart from the coefficients let us next consider the coefficients apart from the literal terms. Take again the derivation of $(a + b)^4$ from $(a + b)^3$:

$$\begin{aligned}(1 + 3 + 3 + 1)(1 + 1) &= 1 + 3 + 3 + 1 \\ &\quad\; 1 + 3 + 3 + 1 \\ &= 1 + 4 + 6 + 4 + 1\end{aligned}$$

This scheme of "detached coefficients" shows clearly that the second coefficient (4) in the expansion of $(a + b)^4$ is reached by adding to the second coefficient (3) in the expansion of $(a + b)^3$ the first coefficient (1); that the third coefficient (6) in $(a + b)^4$ is formed by adding the third coefficient (3) of $(a + b)^3$ to the second (3), and so on. Similarly to find the coefficients in $(a + b)^5$ we have

$$\begin{aligned}(1 + 4 + 6 + 4 + 1)(1 + 1) &= 1 + 4 + \;\; 6 + \;\; 4 + 1 \\ &\quad\; 1 + \;\; 4 + \;\; 6 + 4 + 1 \\ &= 1 + 5 + 10 + 10 + 5 + 1\end{aligned}$$

That is, any coefficient in $(a + b)^5$ is found by taking the corresponding coefficient in $(a + b)^4$ and adding to it its predecessor. It is obvious that the rule must hold good at each successive multiplication. It is also clear that it will always yield unity as the coefficient both of the first and of the last term.

The law of derivation of the coefficients is best exhibited by the following arrangement which has been known to the Chinese since about 1300 but appears to have been first published in Europe by the German algebraist Stifel (1544):—

1					
1	1				
1	2	1			
1	3	3	1		
1	4	6	4	1	
1	5	10	10	5	1
		etc.	etc.		

Starting with the second, the successive rows give the coefficients in the expansions of $(a + b)$, $(a + b)^2$, $(a + b)^3$, etc. The numbers in one row are to be obtained from those in the row above by the rule just discovered. For example, the second 10 in the last row is derived from 4, the corresponding number in the row above, by adding to it its predecessor, 6. The table can, of course, be continued without limit.

§ 2. *Derived Results.*—Stifel's numbers make it possible to write down very easily the results of other expansions. Take the expansion of $(a - b)^5$ as a first example. Here, since $a - b = a + (-b)$, we can derive the expansion we seek from that of $(a + b)^5$ by simply substituting $-b$ for b wherever the latter occurs. Taking the coefficients from the table we have

$$(a - b)^5 = a^5 + 5a^4(-b) + 10a^3(-b)^2 + 10a^2(-b)^3 + 5a(-b)^4 + (-b)^5$$
$$= a^5 - 5a^4b + 10a^3b^2 - 10a^2b^3 + 5ab^4 - b^5$$

Take as a second expansion $(2a + 3b)^4$. Here $2a$ must be put for a and $3b$ for b wherever they occur. Then

$$(2a + 3b)^4 = (2a)^4 + 4(2a)^3(3b) + 6(2a)^2(3b)^2 + 4(2a)(3b)^3 + (3b)^4$$
$$= 16a^4 + 96a^3b + 216a^2b^2 + 216ab^3 + 81b^4$$

The labour-saving virtue of this method is obviously very considerable.

[Ex. XXXI, D, can now be worked.]

CHAPTER XXI.

POSITIVE AND NEGATIVE INDICES.

A. The Uses and Laws of Positive Indices (ch. XVII., § 8; Ex. XXXII).

§ 1. *The Index Notation for Numbers.*—For many purposes "round" numbers are more useful than exact numbers. For instance, we are told that when the census was taken at the stroke of midnight on 5th April, 1911, the population of London was 4,521,685. Supposing this number to have been correct as Big Ben began to strike the hour it is possible that it was incorrect by the twelfth stroke. It was almost certainly wrong by breakfast time and must be some thousands wrong now. In two or three years' time it will be safe to say only that the population is $4\frac{1}{2}$ millions—ignoring the lower denominations. Statisticians in dealing with very large numbers—such as the number of bushels of wheat imported in a year, etc.—generally adopt this plan; they give round numbers. For example, a table of Imports in 1908 in the *Dictionary of Statistics* gives against the entry "Eggs" the number 2,185; against "Iron and Steel" the number 1,119. But the former is a number of millions, the latter a number of thousands of tons. The column headed "value of imports" gives respectively 7,183 and 12,235 against the same names, but each of these is a number of thousands of pounds.

In all such cases the unit must of course be stated in some form. In statistics the statement usually takes the form of a note that the unit is 100,000 cwts. or £100,000, etc. Engineers, physicists, astronomers and mathematicians generally adopt another form—fearing the inconvenience and risk of error involved in changing the units. They get rid of superfluous zeros by what is called the *index notation.* Thus the present population of London can be expressed roundly as $4{\cdot}5 \times 10^6$. The pages of a reference book of Physics or an

Engineer's Pocket Book give many examples of the use of this notation. For example, according to the latest determination, the velocity of light is $3{\cdot}002 \times 10^{10}$—or, less exactly, 3×10^{10}—centimetres per second, the elasticity of steel is about 2×10^{12}. Here are two numbers, constantly written and used, in which the compactness of the index notation is obviously very serviceable. Written in full the latter number would be 2,000,000,000,000.

§ 2. *Multiplication and Division in the Index Notation.*—The advantages of the index notation become still clearer when large round numbers are to be multiplied or divided. Here is an instance. We are told that the nearest fixed star is so far away that light takes about two and a half years to reach it. What is the distance in round numbers? A day has $24 \times 60 \times 60 = 8{\cdot}6 \times 10^4$ seconds, roughly, and in two and a half years there are roughly 9×10^2 days. Hence

$$\begin{aligned} \text{Distance} &= 3 \times 10^{10} \times 8{\cdot}6 \times 10^4 \times 9 \times 10^2 \\ &= 3 \times 8{\cdot}6 \times 9 \times 10^{10} \times 10^4 \times 10^2 \\ &= 232 \times 10^{16} \text{ cms. roughly.} \end{aligned}$$

The rearrangement in the second line is easily justified: to alter the order in which the multiplications are made cannot alter the product (Law of Commutation). The step from "$\times 10^{10} \times 10^4 \times 10^2$" to "$\times 10^{16}$" is also obvious, for both expressions mean that what precedes is to be multiplied by 10 sixteen times.

The centimetre, though the common linear unit of science, is out of place in measuring the distance of a star. If we take the kilometre as unit we must divide by $1 \times 100 \times 1000 = 1 \times 10^5$. Or, taking a mile as, roughly, 1·6 kilometres, we can obtain the answer in miles by dividing by $1{\cdot}6 \times 10^5$ or 16×10^4. The answer is

$$\begin{aligned} \text{Distance} &= \frac{232 \times 10^{16}}{16 \times 10^4} \\ &= 14{\cdot}5 \times 10^{12} \\ &= 145 \times 10^{11} \text{ miles.} \end{aligned}$$

Here again, the procedure is clearly correct. For

$$\frac{232 \times 10^{16}}{16 \times 10^4} = 14{\cdot}5 \times \frac{10^{16}}{10^4}$$

Since the double operation "$\times \dfrac{10^{16}}{10^4}$" means "multiply by

10 sixteen times and divide the result by 10 four times," it may obviously be replaced by the single operation " $\times 10^{12}$ ".

§ 3. *The Law of Indices.*—The instances of § 2 exemplify strikingly the useful properties of the index symbolism, but the properties are, of course, by no means peculiar to the number 10. Let a, b, c, d, and p be any numbers, integral or fractional, directed or non-directed, and let m, n, r, s be any *non-directed whole* numbers. Let $A = a \times p^m$, $B = b \times p^n$, $C = c \times p^r$, $D = d \times p^s$ and let $M = A \times B \times C \times D$. Then we have

$$\begin{aligned} M &= ap^m \times bp^n \times cp^r \times dp \\ &= abcd \times p^m \times p^n \times p^r \times p^s \\ &= abcd \times p^{m+n+r+s} \end{aligned}$$

The justification is simply the former one generalized. The second line follows from the first because the order of multiplication makes no difference to the product. The third follows from the second because " $\times p^{m+n+r+s}$ " merely sums up in one command the successive multiplications by p which " $\times p^m$," " $\times p^n$," etc., order, as it were, in batches.

Similarly if $N = AB/CD$ we have

$$\begin{aligned} N &= \frac{ap^m \times bp^n}{cp^r \times dp^s} \\ &= \frac{ab}{cd} \times \frac{p^m \times p^n}{p^r \times p^s} \\ &= \frac{ab}{cd} \times p^{m+n-r-s} \end{aligned}$$

If after multiplying $(m + n)$ times by p we are to divide $(r + s)$ times by p, we can obviously reach the result more directly by multiplying $(m + n) - (r + s)$ times by p. It is here assumed that the number of multiplications exceeds the number of divisions. If the reverse is the case we must write

$$N = \frac{ab}{cd} \times \frac{1}{p^{r+s-m-n}}.$$

The argument can clearly be extended to any number of numbers, always with the restriction that the indices—because they simply indicate the number of multiplications or divisions—are non-directed whole numbers. Its results can, however, be summed up in the three "laws of indices":—

(i) $\times a^m \times a^n = \times a^{m+n}$

(ii) $\times a^m \div a^n = \times a^{m-n}$ $\quad m > n$

(iii) $\times a^m \div a^n = \div a^{n-m}$ $\quad m < n$

Lastly we may note that although the *laws* of indices deal only with successive multiplication and division, yet indices are commonly used to express single numbers. Thus a^2 and a^3 are very familiar expressions and such symbolisms as $(a + b)^5$ appeared in the last chapter. It is best to regard these forms as convenient ways of expressing $a \times a$, $a \times a^2$, $(a + b) \times (a + b)^4$, etc. Thus we may write

$$\begin{aligned}(a + b)^5 \times (a + b)^7 &= (a + b) \times (a + b)^4 \times (a + b)^7 \\ &= (a + b) \times (a + b)^{11} \\ &= (a + b)^{12}\end{aligned}$$

or more briefly

$$(a + b)^5 \times (a + b)^7 = (a + b)^{12}.$$

[Ex. XXXII can now be worked.]

B. Negative Indices (ch. XVII., § 8; Ex. XXXIII).

§ 1. *Small Numbers.*—The index notation can also be used for expressing compactly the very small numbers with which the chemist, the physicist and the engineer often have to deal. Thus the wave length of yellow light, which is said to be 0·000027 inch, could be written $27 \div 10^6$.

But this notation, though a great improvement upon the extended method of writing small numbers, is not so effective in calculations as it was in the case of large numbers. Consider the following instance. A plate of glass is 0·0081 inch thick. What is the measure of its thickness if the wave length of yellow light is taken as the unit? We have

$$\begin{aligned}\text{No. of wave lengths} &= \frac{81 \div 10^4}{27 \div 10^6} \\ &= \frac{81 \times 10^6}{27 \times 10^4} \\ &= 3 \times 10^2 \\ &= 300\end{aligned}$$

It would be convenient if the step represented by the second line could be avoided.

§ 2. *Negative Indices.*—The following argument suggests a method of doing so :—

$$
\begin{aligned}
2{,}700{,}000 &= 27 \times 10^5 \\
270{,}000 &= 27 \times 10^4 \\
27{,}000 &= 27 \times 10^3 \\
2{,}700 &= 27 \times 10^2 \\
270 &= 27 \times 10^1 \\
27 &= 27 \times 10^0 \\
2{\cdot}7 &= 27 \times 10^{-1} \\
0{\cdot}27 &= 27 \times 10^{-2} \\
0{\cdot}027 &= 27 \times 10^{-3} \\
\text{etc.} &\qquad \text{etc.}
\end{aligned}
$$

We can regard the operation represented by $\times 10^5$ as having the effect of shifting the digits 27 five places to the left; so that the 7 moves from the units' place to the place for hundreds of thousands. Similarly the effect of $\times 10^4$ is to shift the figures four places to the left, and so on. When we multiply by 10 the figures move one place to the left; it is evident, therefore, that 10 can be written 10^1. What are we to say about the next line where the figures are not shifted at all? Obviously, we may indicate this fact by the symbol $\times 10^0$. When we come to 2·7 the figures have moved not to the left but to the right. How shall we indicate this fact? If $\times 10^1$ may be taken as meaning "shift the figures one place to the left" then it is tempting to use $\times 10^{-1}$ to imply "shift them one place to the right". In other words the table suggests that directed numbers may usefully be used instead of non-directed numbers as indices of powers of the number 10; that positive indices may be taken to mean a movement of the digits to the left and so to imply a corresponding number of successive multiplications by 10; and that negative indices may be taken to mean movements of the digits to the right and so to imply a corresponding number of successive divisions by 10. Upon this plan the form "$\times 10^0$" would mean "leave the digits as they are" and so would be equivalent to "$\times 1$".

§ 3. *The Validity of Negative Indices.*—Two questions at once arise: (1) On what principle can negative powers of 10 be combined with one another and with positive powers?—and (2) Is the use of directed indices permissible with numbers other than 10?

The first question is easily answered. Let P and Q be small numbers such that $P = p \div 10^a$ and $Q = q \div 10^b$, a and b being non-directed. Also let $M = PQ$, $N = P/Q$.

Then we can write $P = p \times 10^{-a}$ and $Q = q \times 10^{-b}$, and we have

$$M = p \times 10^{-a} \times q \times 10^{-b} \qquad N = (p \times 10^{-a})/(q \times 10^{-b})$$
$$= pq \times 10^{-a} \times 10^{-b} \qquad = (p \times 10^{-a}) \times (10^{+b} \div q) \quad \text{(i)}$$
$$= pq \times 10^{-(a+b)} \qquad = p/q \times 10^{-a} \times 10^{+b} \quad \text{(ii)}$$
$$= pq \times 10^{[(-a)+(-b)]} \qquad = p/q \times 10^{(-a+b)} \quad \text{(iii)}$$
$$= p/q \times 10^{[(-a)-(-b)]} \quad \text{(iv)}$$

Taking the product first the argument runs as follows: The order of multiplications and divisions is indifferent; hence line (i), which signifies that the product pq is to be divided by 10 a times and then again b times. That is, pq is to be divided by 10 $(a + b)$ times, a fact expressed by line (ii). But $-(a + b) = (-a) + (-b)$; that is, when the total number of divisions by 10 is expressed by a single index, that index is the algebraic sum of the two original indices.

The argument concerning the quotient, N, takes a similar course. Line (i) follows by the fundamental rule for division by a fraction. Line (ii) merely gives the same operations in a different order. Since the complex "$\times 10^{-a} \times 10^{+b}$" means "divide p/q by 10 a times and then multiply by 10 b times," it may be replaced, as in line (iii) by "$\times 10^{-a+b}$". For if a is greater than b this can be written "$\times 10^{-(a-b)}$" and means "divide by 10 $(a - b)$ times"; while if a is less than b it can be written "$\times 10^{b-a}$" and means "multiply by 10 $b - a$ times". One of these alternative instructions must be equivalent to the former pair. Lastly we have that since $-a + b = (-a) - (-b)$ the single index is the algebraic difference of the two original indices.

Precisely similar arguments may be followed in cases in which one index is positive and the other negative or both positive. They will all lead to the same result, namely that the laws of indices discussed in A § 3 may be extended to directed indices of 10 with the substitution of *algebraic* addition and subtraction for *arithmetical* addition and subtraction.

Further, the argument holds good for all numbers. Let m and n be two directed numbers, positive or negative, whose numerical values are a and b. That is let $m = \pm a$ and $n = \pm b$. Let p be any other number, directed or non-directed. Then $\times p^m = \times p^{\pm a}$, and $\times p^n = \times p^{\pm b}$, the symbols implying multiplications by p when the index is

positive, divisions when it is negative. Moreover it is clear that

$$\times\, p^{\pm a} \times p^{\pm b} = \times\, p^{\pm a \pm b}$$

the same sign being taken with a or b on the right as is taken on the left of the sign of equality. For instance

$$\times\, p^{+a} \times p^{-b} = \times\, p^{+a-b}$$

For the left-hand side means a multiplications followed by b divisions; that is, $a - b$ multiplications if $a > b$ and $b - a$ divisions if $a < b$. But either of these operations can be represented by $\times\, p^{a-b}$. For if $a > b$, $a - b$ is positive and the symbols mean $a - b$ multiplications by p. On the other hand if $a < b$, $a - b$ is negative and may be written $-(b - a)$; so that $\times\, p^{a-b} = \times\, p^{-(b-a)}$, which, by hypothesis, means $b - a$ divisions by p.

The other possible combinations of signs can be examined in the same way and the equivalence

$$\times\, p^{\pm a} \times p^{\pm b} = \times\, p^{\pm a \pm b}$$

established in each case. But

$$\pm a \pm b = (\pm a) + (\pm b) = m + n.$$

Hence in all cases

$$\times\, p^{m} \times p^{n} = \times\, p^{m+n}$$

By similar arguments we have that

$$\times\, p^{\pm a} \div p^{\pm b} = \times\, p^{\pm a \mp b} = \div\, p^{\pm b \mp a}$$

For instance, $\times\, p^{-a} \div p^{+b}$ means a divisions by p followed by b more divisions, that is $a + b$ divisions in all. This total operation may be represented by $\div\, p^{a+b}$ or $\times\, p^{-a-b}$. The other cases lead to similar results. But

$$\pm a \mp b = (\pm a) - (\pm b) = m - n, \text{ and}$$
$$\pm b \mp a = (\pm b) - (\pm a) = n - m.$$

Hence we have in all cases

$$\times\, p^{m} \div p^{n} = \times\, p^{m-n} = \div\, p^{n-m}$$

We conclude that powers of any number may be expressed by positive indices to indicate successive multiplications by that number and by negative indices to indicate successive divisions, and that such directed indices may be combined in accordance with the laws followed by non-directed indices.

Finally we may inquire whether the assumption that $p^0 = 1$ will also harmonize with these laws. If $p^0 = 1$ then it is clear that

$$p^n \times p^0 = p^n \times 1 = p^n \quad \text{and} \quad p^n \div p^0 = p^n \div 1 = p^n$$

But in the first case p^n may be written as p^{n+0} and in the second case as p^{n-0} without error. Hence if p^0 be taken to be 1 the symbol may be used in combination with other symbols without the need of any modification of the laws of non-directed indices.

[Exs. XXXIII and XXXIV may now be worked.]

CHAPTER XXII.

ALGEBRAIC DIVISION.

A. Algebraic Division (ch. XVII., § 9; Ex. XXXV).

§ 1. *Nature of Algebraic Division.*—As was seen in ch. XVIII., C, § 5, to divide − 12 by − 3 is to ask by what number − 3 must be multiplied to yield − 12. Similarly to divide (say) $a^3 + b^3$ by $a + b$ is to inquire what is the other factor P such that

$$a^3 + b^3 = (a + b)\mathrm{P}$$

This other factor P is the algebraic *quotient* of $a^3 + b^3$ by $a + b$. As we know, it is $a^2 - ab + b^2$.

The matter can also be put in the following way. The expression $(a^3 + b^3)/(a + b)$ is an algebraic fraction. Now it happens in this case that the numerator can be factorized into $a + b$ and $a^2 - ab + b^2$. Hence

$$\frac{a^3 + b^3}{a + b} = \frac{(a + b)(a^2 - ab + b^2)}{a + b}$$
$$= a^2 - ab + b^2$$

That is, the algebraic fraction $(a^3 + b^3)/(a + b)$ could be replaced in any formula by the expression $a^2 - ab + b^2$ which is not fractional. For this reason $a^2 - ab + b^2$ is called the *integral equivalent* of the algebraic fraction $(a^3 + b^3)/(a + b)$.

If the numerator of an algebraic fraction does not contain the denominator as one of its factors it has of course no integral equivalent. Thus $(a^2 + 1)/(a + 1)$ has no integral equivalent. Nevertheless we can in this case write

$$\frac{a^2 + 1}{a + 1} = \frac{(a + 1)(a - 1) + 2}{a + 1}$$
$$= a - 1 + \frac{2}{a + 1}$$

Here the equivalent of $(a^2 + 1)/(a + 1)$ is partly integral and partly a fraction. But the fraction $2/(a + 1)$ differs from the

original fraction in an important respect. The numerator of $(a^2 + 1)/(a + 1)$ is of a *higher degree* than the denominator as regards the variable a; the numerator of $2/(a + 1)$ is of *lower degree* than the denominator. An algebraic fraction in which the degree of the numerator, N, is, with respect to a variable in the denominator, of higher degree than the denominator, D, may be called an *improper* algebraic fraction. In such a case we can always write N = DP + Q where Q is of a lower degree than D.

§ 2. *The Division Process.*—When the algebraic quotient cannot be seen upon inspection it should be sought in the following way :—

Example: Divide $a^4 - 5a^3 + 13a^2 - 29a + 24$ by $a - 3$. Solution :—

$$a^4 - 5a^3 + 13a^2 - 29a + 24 = (a - 3)(a^3 - 2a^2 + 7a - 8).$$

Our object is to find the second factor or, failing a complete quotient, to express the dividend in the form $(a - 3)P + Q$. The first term of the factor must be a^3. Multiplied by -3 this would give a term $-3a^3$; the actual term in the dividend involving a^3 is $-5a^3$, hence we must provide for a term $-2a^3$. This is done by adding to the second factor the term $-2a^2$ which gives $-2a^3$ when multiplied by a. But when multiplied by -3 the $-2a^2$ gives $+6a^2$ while the dividend contains $+13a^2$. We need, therefore, a term $+7a^2$. To produce this we must add to the new factor a term $+7a$. This term multiplied by -3 gives $-21a$, instead of the $-29a$ of the dividend, and so on.

With a little practice division by a trinomial can be carried out in the same way. At first, however, it is best to proceed as follows :—

Divide $6a^4 + 7a^3 - 15a^2 + 47a - 21$ by $2a^2 + 5a - 3$ Solution :—

$$\begin{aligned} 6a^4 + 7a^3 - 15a^2 + 47a - 21 &= (2a^2 + 5a - 3)(3a^2) - 8a^3 - 6a^2 \\ &\quad + 47a - 21 \\ &= (2a^2 + 5a - 3)(3a^2 - 4a) + 14a^2 \\ &\quad + 35a - 21 \\ &= (2a^2 + 5a - 3)(3a^2 - 4a + 7) \end{aligned}$$

The first term of the quotient must be $3a^2$. But $+5a \times 3a^2$ gives $+15a^3$ instead of $+7a^3$ and $-3 \times 3a^2$ gives $-9a^2$ instead of $-15a^2$. We set down, therefore, the complement $-8a^3 - 6a^2 + 47a - 21$ which is necessary to make the right-hand expression equivalent to the left. The term $-8a^3$

may be removed from this complement by adding $-4a$ to the $3a^2$. The complement now assumes the value $14a^2 + 35a - 21$. The next stage removes it altogether.

B. Geometric Series (ch. XVII., § 9; Ex. XXXVI).

§ 1. *A Fraction as the Sum of a Series.*—A long defunct London newspaper, *The News*, published in its issue of January 10, 1813, the following piece of intelligence: "A few days since a bargain was made at Oswestry market between a farmer whose name is Evans and the ostler at the Crosskeys Inn, for a goose which weighed 11 lb. The ostler agreed to give Mr. E. one halfpenny for the first pound, a penny for the second and in like manner to double the sum for every succeeding pound, which raised the price of the goose to . . ." But instead of hearing at once the conclusion of the matter it will be more interesting if—no doubt like the astute Mr. Evans—we determine the cost of the goose by calculation.

If P is the cost in halfpence we have

$$\begin{aligned} P &= 1 + 2 + 4 + 8 + \ldots \\ &= 1 + 2 + 2^2 + 2^3 + \ldots + 2^{10} \end{aligned}$$

The calculation of the sum of this series can be greatly abbreviated by noting that it is $1 + a + a^2 + \ldots + a^{n-1}$ with $a = 2$, $n = 11$. But by Ex. XXXV, No. 5, we have

$$\begin{aligned} 1 + a + a^2 + \ldots + a^{n-1} &= \frac{1 - a^n}{1 - a} \\ &= \frac{a^n - 1}{a - 1} \end{aligned}$$

Hence in this case

$$\begin{aligned} P &= (2^{11} - 1) \text{ halfpence} \\ &= 2047 \text{ halfpence} \\ &= £4\ 5s.\ 3\tfrac{1}{2}d. \end{aligned}$$

We are told that the discovery of this answer to the problem caused "great mortification to the purchaser".[1] No wonder!

Here is another problem. A tree grows 1 foot during the first year of its life; in each successive year its increase in height is nine-tenths of the increase during the previous year. How high will it become in a given number of years?

[1] The story was reprinted in the London *Observer* of January 13, 1913, under the heading "A Hundred Years Ago".

In this case we have for the height h after n years

$$h = 1 + 0{\cdot}9 + (0{\cdot}9)^2 + (0{\cdot}9)^3 + (0{\cdot}9)^4 + \ldots + (0{\cdot}9)^{n-1}$$
$$= \frac{1 - (0{\cdot}9)^n}{1 - 0{\cdot}9}$$
$$= 10\{1 - (0{\cdot}9)^n\}$$

Now the interesting thing about this answer is that the term $(0{\cdot}9)^n$ or $\{\frac{9}{10}\}^n$ becomes smaller as n increases. For whatever value it has for a given year its value for the next year will be only nine-tenths of the previous value. Thus, in time, $(0{\cdot}9)^n$ will become too small to be measurable or visible—smaller, in fact, than any minutest fraction of a foot that you can name. In other words, the second factor of the expression $10\{1 - (0{\cdot}9)^n\}$ can be made to differ from 1 by as little as we please. Hence the height of the tree will in time differ from 10 feet by an entirely inappreciable amount. For all practical purposes, then, 10 feet is the final height which the tree would reach even if it grew for ever. This fact is conveniently expressed by saying that the sum of the series which gives us the height of the tree *approaches* 10 as the number of terms increases. The full meaning of this statement is that, although the tree will never reach exactly 10 feet, yet if you name a height as little short of 10 feet as you please the tree's height will in time become still nearer to 10 feet, and will ever after remain still nearer. No such statement could be made about a series of increasing terms.

§ 2. *Geometric Series.*—These examples suggest a method that can be used for summing any series in which each term is obtained by multiplying its predecessor by a constant factor. Such a series of n terms can always be expressed in the form

$$a + ar + ar^2 + \ldots + ar^{n-1},$$

r being the constant factor or (as it is generally termed) constant *ratio*. Then we have for the sum S:—

$$S = a + ar + ar^2 + ar^3 + \ldots + ar^{n-1}$$
$$= a(1 + r + r^2 + \ldots + r^{n-1})$$
$$= a \cdot \frac{1 - r^n}{1 - r}$$
$$= a \cdot \frac{r^n - 1}{r - 1}.$$

The last form of the sum is more appropriate when r is

numerically greater than 1 and the earlier form when r is numerically less than 1.

Series of this kind are called *geometric series.* (See Note on Ex. XXXVI, No. 6.)

In general we have, by Ex. XXXV, No. 24, that

$$\frac{a}{1-r} = a\left(1 + r + r^2 + \ldots + r^{n-1} + \frac{r^n}{1-r}\right)$$

$$= a + ar + ar^2 + \ldots + ar^{n-1} + \frac{ar^n}{1-r}$$

Now if r is numerically less than 1 the fractional complement $\frac{ar^n}{1-r}$ becomes smaller and smaller as n increases and by taking n large enough may be made smaller than any number that any one chooses to specify. Hence the fraction $a/(1-r)$ becomes ever more approximately equal to the sum of the series the more numerous the terms. When the terms are unlimited in number or are very numerous it may be taken, therefore, as the sum of the series. Since, however, in strictness the complement $ar^n/(1-r)$ always retains some value, though a negligible one, $a/(1-r)$ is called the "sum to infinity". This term simply means that by a convenient fiction we may suppose that if the terms were "infinite" in number the sum would be exactly $a/(1-r)$. It should be remembered that this is nothing *more* than a convenient fiction. The actual facts are, as stated, that by making the terms numerous enough the exact sum of the series may be brought as near to $a/(1-r)$ as we please, and that the addition of further terms will bring it still nearer to this number.

If r is positive and <1 the complement $ar^n/(1-r)$ is always positive. The fraction $a/(1-r)$ marks, therefore, a limit which the sum of the series never quite reaches. But if r is negative odd powers of r are negative and the corresponding values of the complement $ar^n/(1-r)$ will be negative, while those corresponding to even values of n will as before be positive. Thus the sum of the series will be alternately greater and less than $a/(1-r)$ but the successive values of the sum will swing less and less above and below $a/(1-r)$ as n increases. The table below gives the values of the sum of

$1 - \frac{1}{2} + \frac{1}{4} - \frac{1}{8} + \frac{1}{16} - \ldots$ for 1, 2, 3, . . . 8 terms.

Here the "sum to infinity" is

$$S = \frac{1}{1 + \frac{1}{2}}$$
$$= \frac{2}{3}$$
$$= 0{\cdot}6666 \ldots$$

and it will be seen that the odd terms approach this value from above and the even terms from below. However far the series is continued no term will exactly reach it.

$n =$	1		3		5		7	
$S =$	1		0·75		0·6875		0·671875	
$S =$		0·5		0·625		0·65625		0·6640625
$n =$		2		4		6		8

[Ex. XXXVI may now be worked.]

CHAPTER XXIII.

THE COMPLETE NUMBER-SCALE.

(Ch. XVII., § 10; Exs. XXXVII, XXXVIII.)

§ 1. *The Complete Number-Scale.*—It will be well at this point to summarize and complete the chief ideas which underlie the work of Exs. XXVII-XXXVI.

In Ex. XXVII we began to face problems the study of which is greatly facilitated by the device of attaching *plus* and *minus* signs to the numbers of arithmetic. In that way we reached the notion of a scale of directed numbers (symbolized by the points of an endless line AB) which may be supposed to start at zero and be continued without end each way, so that a mirror held across the series at the zero-point, O, and facing A would give OB, the negative part of the scale, as a reflection of OA, the positive part. One important advantage of using this scale is that when we are dealing with any set of things which naturally fall into a settled order with regard to one another we can assign a special number to each of them and need have no fear that the assignment will be upset by the occurrence of new members of the series for which provision was not originally made. Thus when Fahrenheit adopted (1714) the thermometer graduation which is commonly used in England, he chose 32° below freezing-point as the starting-point of his scale, because he believed that this graduation represented the greatest cold obtainable. It was, of course, soon found that even the natural coldness of the air of the Arctic regions is often more intense than the coldness which brings the mercury to the Fahrenheit zero. But by giving signs to the degrees and continuing the scale below zero as a negative graduation a means was easily found of placing records of Arctic cold in the same series as the records of our own more moderate climate.

A greater advantage is that by the use of directed numbers

problems which would have to be considered separately if we were restricted to the ordinary numbers of arithmetic can all be studied together and solved by a single algebraic investigation. Take as a simple instance the formula which gives the distance (d) apart after time t of two motor cars which move along the same road with different speeds (v_1, v_2) and are originally at a given distance apart (d_0). If our symbols stood only for non-directed numbers we should have to give a whole set of formulæ to suit the various cases—cars going the same way, the faster in front; going the same way the slower in front; going different ways and towards one another; different ways but away from one another, etc., etc. But with the aid of directed numbers all possible cases are included in the single formula $d = d_0 + (v_1 - v_2)\ t$. This advantage depends entirely upon the fact that all possible velocities and distances both to the right and to the left, and all possible times both in the future and in the past can be treated respectively as members of a single series of velocities, distances or times by assigning directed numbers to them. Exs. XXVIII and XXIX contain many instances of this labour-saving virtue of directed numbers.

One more example. Suppose there are a number of persons possessed of different sums of ready money—£1, £5, £10, £100, etc.,—and a number of others who possess nothing but owe different sums—£1, £5, £20, etc. These can all be arranged in one series in order of their financial position. If we take zero to represent the financial position of the person who has empty pockets but no debts then the position of A who possesses £5 can be labelled + 5 and that of B who owes £5 labelled − 5, etc. Upon this plan transactions which carry a person from the category of possessors into the category of debtors can be brought into line with those which affect merely the amount of possessions or debts. Thus if A incurs a liability for £20 his financial position is carried from + 5 to − 15.

It is often said of directed numbers that (for example) while + 20 is greater than + 18, − 20 is less than − 18 and that all negative numbers are less than nothing. The last example shows the inadvisability of such a statement. A person who owes £20 does not possess less than a person who owes £18, for neither possesses anything, and, indeed, it is obviously absurd to suppose anything to be less than

nothing. Directed numbers should be spoken of as differing not in magnitude but in position with regard to the zero. Perhaps the most convenient form of statement to use is that, of two numbers on the scale, the first, represented in fig. 63 by a point to the right of the point which represents the second, is *higher* in the scale, and the second number *lower* in the scale. Thus − 18 is higher than − 20 but + 18 lower than + 20, while all positive numbers are higher than all negative numbers and all negative numbers are lower than zero. When a and b are directed numbers the symbolisms $a > b$, $a < b$ should always be read "a is higher than b," "a is lower than b".

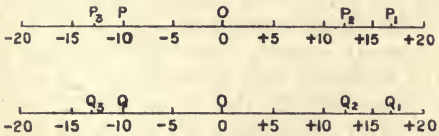


FIG. 63.

§ 2. *Operations on Directed Numbers.*—Let the points P_1 and P_2 (fig. 63) represent two directed numbers a and b. Then we may think of all operations performed on a and b as having the effect of moving P_1 and P_2 higher or lower along the scale. Thus $a \times n$, if n is a non-directed number such as 3, or 4·2, or 3/11, simply carries P_1 to a point n times OP_1 from O but can never cause it to pass over O. The result of $a + b$ could be shown by sliding the segment OP_2 to right or left until the end now at O coincides with P_1. P_2 then marks the point to which P_1 must be shifted if it is to represent $a + b$. The operation may or may not carry P_1 across O. The result of $a - b$ is shown by sliding OP_2 until P_2 coincides with P_1. The other end of the segment now marks the point to which P_1 must be shifted to represent $a - b$. For the movement from O to this point represents the combination $a + (-b) = a - b$.

The effect of multiplying a by a *directed* number b is best thought of in the first instance as the production of a rectangle on the positive (upper) or negative (lower) side of OP_1. But since the number of positive or negative units of area in the rectangle will have a place at some point R on AB the effect may finally be taken to be to shift P_1 to R. P_1 will thus

be carried across O if b is negative and left farther from or nearer to O according as P_2 is without or within the range from + 1 to − 1. If b is negative and $-1 < b < +1$, successive multiplication by b (i.e. ab^n, n positive) will carry P_1 backwards and forwards across O and bring it constantly nearer to O—but never quite up to it. The effects of single or repeated division of a by b (i.e. ab^n, n negative) can be investigated similarly.

There is only one kind of operation that gives rise to difficulty. Wherever P_1 is originally, the operation a^2 always leaves it on the positive side of O. It follows that the inverse operation $\sqrt{a}$ can be performed only if P_1 starts on the positive side. In other words we cannot take the square root of a negative number. This difficulty does not arise with an odd power such as a^3, for $a \times a^2$ will leave P_1 on the positive side of O if it starts on that side and on the negative side if it starts there. Moreover, since $a \times a$ always leaves P_1 on the positive side no matter on which side it started the result of $\sqrt{a}$ is ambiguous. It may imply carrying P_1 on to the negative side of O or leaving it on the positive side. That is, every positive number has two square roots—one positive and the other negative.

A further difficulty which may affect all root-operations is that of "surds". As we saw (ch. VII., B, § 2) there is really no number on our scale which when squared gives (for example) + 20. We could by a geometrical construction define a point P so distant from O that the area of the square upon OP is + 20, but we cannot name any whole or fractional number which belongs to that point. We can only find *pairs* of numbers, belonging to points that constantly approach each other, between which the unnumbered point lies. For purposes of calculation this solution of the difficulty suffices. If we want anything better we must regard "$\sqrt{20}$" as itself the label of this point in our graduated line.

§ 3. *Fractions.*—We have seen that combinations of directed numbers may be treated according to the same rules as combinations of non-directed numbers connected by *plus* and *minus* signs bearing the ordinary arithmetical senses of "add" and "subtract". The manipulation of fractions in which the numerators and denominators are directed numbers has not yet been formally examined.

In ch. IX. it was assumed for convenience that an algebraic fraction, such as a/b, always describes an arithmetical fraction. This need not be the case. For example, a/b may measure the ratio of the distance (a) that a point has travelled to the time (b) that the journey has taken. In that case if the distance (say – 12 feet) is an exact multiple of the time (say + 3 seconds) the number represented by a/b will be integral (– 4). Thus the value of a/b when a and b are directed may be represented by *any* point on our scale. That is, although a/b is always fractional in *form* it need not be fractional in *value*.

We must now inquire whether algebraic fractions whose numerators and denominators are directed numbers can be treated in the same way as the fractions of Exs. XII and XIII. To begin with, can we say that $\frac{a}{b} \times \frac{c}{d} = \frac{ac}{bd}$? Let us suppose that a/b (represented by P, fig. 63) is negative and c/d (represented by P_1) is positive. Then a and b must have different signs and c and d the same sign. The various possibilities, so far as they affect the sign of the product, may be represented thus :—

$$\text{(i)}\ \frac{+}{-} \times \frac{+}{+} = \frac{+}{-} \qquad \text{(ii)}\ \frac{-}{+} \times \frac{+}{+} = \frac{-}{+}$$
$$= -\qquad\qquad = -$$
$$\text{(iii)}\ \frac{+}{-} \times \frac{-}{-} = \frac{-}{+} \qquad \text{(iv)}\ \frac{-}{+} \times \frac{-}{-} = \frac{+}{-}$$
$$= -\qquad\qquad = -$$

Thus, however the operations may be supposed to occur, the final result will always be to bring the indicating point to the same position (say R) on the negative side of O. Similar investigations, assuming that a/b and c/d are both positive or both negative, obviously lead to the same result. We conclude that we may write $\frac{a}{b} \times \frac{c}{d} = \frac{ac}{bd}$ in all cases, without inquiry into the signs of the numbers.

Next, is the value of a/b affected by multiplying numerator and denominator by the same (directed) number d? Obviously not, for (by the preceding argument)

$$\frac{ad}{bd} = \frac{a}{b} \times \frac{d}{d} = \frac{a}{b} \times (+ 1).$$

It follows that the operation of "reduction to a common

denominator " for the purpose of adding or subtracting fractions may be carried out with directed as with non-directed numbers. The only difference is that the addition and subtraction are themselves algebraic.

Finally, since we may treat bracketed groups of directed numbers just as we may treat similar groups of non-directed numbers, *all* cases of the " simplification " of algebraic fractions may be performed with symbols representing directed numbers exactly as with those representing non-directed numbers.

§ 4. *Equations.*—A last question remains. Do the rules for " changing the subject " hold good in the case of relations between directed numbers? As an example consider the problem of finding the value of n from

$$a + \frac{b}{q - pn} = c$$

all the symbols standing for positive or negative numbers.

If we knew the value of n and substituted it on the left the numbers on the two sides of the relation would, of course, be identical. Let the point P moving along the upper scale (fig. 63) record the values of the left-hand side and Q moving along the lower scale those of the right-hand side. Then P and Q occupy, at first, corresponding positions on their respective scales. To remove a from the left add the component $(-a)$ to both sides. This operation must have the same effect both on P and Q, moving them (say) to P_1 and Q_1. The relation becomes, therefore,

$$\frac{b}{q - pn} = c - a.$$

Next multiply both sides by $q - pn$. Here again since the same operation is performed upon numbers occupying the same position in the number scale the results must be identical. We may suppose them to be represented by P_2 and Q_2. We may write, therefore,

$$b = (c - a)\,(q - pn).$$

Now divide each side by $(c - a)$. By the same argument the results, represented by P_3 and Q_3, must be identical and we may write

$$\frac{b}{c - a} = q - pn \quad . \quad . \quad . \quad \text{(i)}$$

By continuing this process—first adding $-q$ to both sides

and then dividing by $-p$—we bring the representative points to their final positions—still identical—and we have

$$n = \{b/(c - a) - q\}/(-p)$$
$$= \{q - b/(c - a)\}/p$$

In this example all the typical operations involved in changing the subject have been involved, and the argument shows that, although they must be justified on different grounds from those of ch. x., yet the operations themselves may be carried out in exactly the same way upon directed as upon non-directed numbers.

Suppose that when the calculation had reached the stage represented by (i) and by the points P_3 and Q_3, a component $-(q - pn)$ had been added to each side. The obvious effect on Q would be to bring it to zero. P would therefore have come to zero also. The relation would then have read

$$b/(c - a) - (q - pn) = 0$$
$$\text{or} \quad pn + b/(c - a) - q = 0 . \quad . \quad \text{(ii)}$$

No matter what the original relation may be it can always be reduced, by the algebraic addition to each side of a suitable component, to the form in which the right-hand side is zero. This is to be regarded, therefore, as a standard form of expression. It plays so important a part in algebra that we shall give it a special name—*equation*. It is true that this term is often applied to relations in other forms, such as (i), but there is an advantage in reserving it for the standard form and we shall generally follow that practice. To find the value of a variable from a relation expressed in this standard form is called "solving the equation" and any value of the variable which satisfies the relation is called a *root* of the equation.

[Exs. XXXVII and XXXVIII may now be worked.]

CHAPTER XXIV.

THE PROGRAMME OF SECTION II (EXS. XXXIX-L)

§ 1. *The Contents of the Group.*—It has already been said (ch. v., § 3) that the second group of exercises of Section II corresponds in character to the second group of Section I. Thus those exercises, like these, fall naturally into two sub-divisions. In Exs. XXXIX and XLII-XLIX the simple study of proportionality of the earlier section is, by the introduction of directed numbers, expanded and generalized into an elementary doctrine of functions of one variable of the first and second degrees, culminating in a study of the relations between these functions which are covered by the technical terms "differentiation" and "integration". In Exs. XL and XLI directed numbers are used in a similar way to extend the range of the trigonometrical ratios to all angles up to 360°. In Ex. L this part of the work is completed by a simple inquiry into the methods by which tables of the trigonometrical ratios may be calculated. It will be more convenient to consider the exercises of the sub-divisions separately than to discuss the contents of the group *seriatim*. The first half of this chapter is, accordingly, given to the exercises in which the idea of a function is developed, and the second half to the trigonometrical exercises.

§ 2. *Linear Functions* (Ex. XXXIX, ch. xxv., A).—The discussion in ch. xviii., C, and the graphical examples in Ex. XXIX have prepared the way for ch. xxv., A, and the corresponding exercise. The relation of the present argument to the work upon which it is based is that described in ch. xi., § 1, and ch. xii., § 4. That is, from the study of concrete cases in which the variables are connected by a linear relation we are now to turn our attention to the relation itself. As before, our study of the relation is to be guided largely by consideration of the familiar properties of its

"graphic symbol"—the straight line—while y and x are to be our algebraic symbols for the words "dependent variable" and "independent variable".

The lesson is simple but the teacher should attend carefully to a point of logic which is sometimes slurred. The original definition of the tangent of an angle (ch. XIII., A) applies only to angles less than 90°, for it is based upon the properties of right-angled triangles. If, then, angles between 90° and 180° are to be considered as possessing tangents these must be defined in a new way. For the present the simplest way is that of ch. XXV., A, § 2; *by definition*, the tangent of an obtuse angle is to be considered equal to that of its supplement but negative in sign.

In connexion with Ex. XXXIX the teacher should read again ch. XI., § 1. All the examples must be considered important because of their direct bearing upon much subsequent work. The teacher is advised to pay particular attention to the device of "shifting the graph" which will be constantly used throughout the rest of the book. This device is equivalent analytically to the more usual "change of origin," but is preferred here as a more vivid, attractive and intelligible idea. In demonstrating it upon the squared blackboard the teacher may conveniently make use of a piece of thin white tape to be secured by a couple of drawing pins, or a metre rule turned edgewise to the class.

In **No. 3** the positions of the lines should be determined by substituting first $y = 0$ and then $x = 0$, and so finding where they cross the axes.

In **No. 4** (i) the answer is, of course, that when expressed in the form

$$y = ax + b$$

a is the same for both lines. In No. 4 (ii) the coefficients of x are respectively $+\frac{3}{4}$ and $-\frac{4}{3}$. The angle whose tangent is $\frac{4}{3}$ is the complement of the angle whose tangent is $\frac{3}{4}$. Calling the former α, the latter is $90° - \alpha$. It follows that the angle whose tangent is $-\frac{4}{3}$ is $180° - (90° - \alpha)$, i.e. $90° + \alpha$; that is, the two lines are perpendicular. The same argument is generalized in **No. 5.**

§ 3. *Hyperbolic and Parabolic Functions* (Ex. XLII, ch. XXVI., A, B).—The arguments of these lessons and examples are the natural extension of those of ch. XVI., A, B, and Exs. XXIII, XXIV. The exploration of the "fields" of

the two kinds of functions offers an excellent opportunity for heuristic work on the part of the class, and an admirable illustration of the power of a graphic method to stimulate and guide mathematical thinking. The teacher should, in particular, refrain from anticipating the class's discovery of the situation of the branches needed to complete the hyperbola and the parabola, and should be ready to make good use of inevitable first mistakes.

The examples showing how the position of the graph depends upon the form of its algebraic expression are of great importance and should be studied by means of a movable graph drawn upon tracing paper (**Nos. 1-4, 9-11**). A movable parabola for blackboard demonstration is easily made by shaping a length of stiff iron or copper wire, and may be used with much effect. The construction of a movable hyperbola is not so simple since it must include some device for securing to one another the two branches of the curve. In any case Nos. 3, 4, 10, 11, 13 should be taken orally, the movements of the graphs being made simultaneously by the whole class.

Nos. 15-18 should be emphasized since they contain the essence of the method which is to be used in Ex. XLIII for the solution of quadratic equations.

Parabolic functions involving positive and negative values of the variable are of great importance in the physical world. For example, the path followed by the centre of a cricket ball thrown into the air is very nearly a parabola; that is, its vertical height is a parabolic function of its horizontal displacement. The path of a bullet fired from a rifle is less perfectly parabolic because it is distorted by a greater air-resistance.

The practical study of one or two cases of parabolic motion is a legitimate application of the methods of Ex. XLIII and will do much to secure mastery of the theory of the function. Mr. G. Goodwill (*Elementary Mechanics*, Clarendon Press, 1913) has described (pp. 34-6) some beautiful experiments for tracing the path of a projectile. **Nos. 29, 30** of Ex. XLIII indicate another more easily performed. For the purposes of the mathematical classroom the following extremely simple method will be found quite satisfactory (**No. 28**). Pin a sheet of squared or plain paper on a smooth drawing-board. Fix the board in a sloping position with its lower edge at the edge of the table. Take a smooth and uniform

ball (a motor-car ball-bearing or a billiard ball has the requisite smoothness and uniformity) and smear it lightly with lubricating oil. After one or two trials (performed before the oiling) it will be found easy to project the ball diagonally across the drawing-board so that it leaves an oily parabolic track. A line can be drawn to preserve the record which can be given out to the class for study.

§ 4. *Quadratic Equations* (Ex. XLIII, ch. XXVI., C).—Quadratic equations are undoubtedly important but they have loomed too large in the elementary algebra course. They are too often introduced without an adequate motive—generally, perhaps, because after "doing" equations with one root it is time to pass on to equations with two! Moreover, as the result of this formal and unpedagogical mode of treatment, they are too often nothing but formal exercises leading to "problems" more than usually unconvincing and infertile. All these disadvantages are avoided by treating the study of quadratics as merely an episode in the general elementary theory of parabolic functions. This is the point of view adopted in ch. XXVI., C, and Ex. XLIII. The pupil has already learnt a good deal about the parabolic function and its graph—including the determination of turning values and the turning points which correspond to them. He knows that, in general, a given value of the function is produced by two different values of x; it is natural, therefore, to inquire how these values may be calculated. In the search for an universal method of dealing with this problem he discovers that a particular case—the case of $y = 0$—can be solved (when it is soluble at all) by the simple device of factorizing the function, and finds also that all other cases can be reduced to this case. Thus the determination of the values of x for which the value of the function vanishes is seen to be a process of considerable technical importance and therefore worthy of detailed study.

This way of approaching the subject leads to a noteworthy consequence, namely, that no attention is given to the so-called "imaginary roots". In the case of every parabolic function the field of y has a certain definite upper or lower limit. Each value of y within the field corresponds to two different values of x with the exception of the value of y which bounds the field. This last corresponds, strictly speaking, to only one value of x, though it is convenient to adopt the convention that x has in this case two identical values. To ask

what values of x correspond to a value of y outside the field of y is to ask a self-contradictory and, therefore, absurd question. It follows that some quadratic equations will have two roots, others one root—or, if you prefer to put it so, two identical roots—while others will have no roots at all.

There should be no need to defend this way of regarding the matter from the standpoint either of teaching practice or of mathematical theory. The "imaginary" root is certainly a rock of stumbling to the thoughtful beginner whose common sense is offended by the supposition that a negative number may in any sense possess a square root. On the other hand, he is not at all perturbed to find that some equations have no solutions—especially when he sees that this circumstance is correlated with an obvious property of the parabolic graph—and is, as a rule, secretly of opinion that it is "silly" to insist upon finding them. Again, the notion of "imaginary" numbers is equally offensive to modern mathematical theory. It is, in fact, simply another remnant of the cloud of confusion that shrouded the early history of many of the main ideas of algebra (cf. ch. XVII., § 1). It is high time to delete the term from elementary text-books and to get rid of its misleading associations. But this act of salutary purgation demands as its correlative a suitable treatment of the doctrine of "complex numbers" which replaces in modern theory the illegitimate idea of imaginaries. No doubt it would be possible to make the study of quadratics the occasion for teaching the elements of this doctrine, but there is much more to be said for postponing it to a later stage in mathematical instruction. Postponement both avoids the risk of blurring the clear lines of association between the quadratic equation and the parabolic function, and secures that the student shall attack the question of complex numbers with a mind mature enough to appreciate the logical subtleties involved in it. For these reasons the suggestion that quadratic equations which are apparently insoluble may after all be solved is reserved until the student is well into Part II of this work. The inevitable consequence that many boys and girls may leave school without hearing that *all* quadratic equations have roots is one which we ought, perhaps, to be able to face with equanimity.

Most of the examples in Ex. XLIII are of familiar types. The solution of **No. 16** is, of course,

$$(x - m\alpha)(x - m\beta) = 0$$

or

$$x^2 - m(\alpha + \beta)x + m^2\alpha\beta = 0.$$

It shows that to produce an equation whose roots are m times those of a given equation we need only multiply the coefficient of x by m and the constant term by m^2. Applying this principle to **No. 17** (i) we have for the required equation

$$x^2 - 8x - 48 = 0.$$

No. 18 illustrates a rather useful application of this method. In (i) we have

$$7x^2 + 4x - 3 = 0$$

whence $$x^2 + \tfrac{4}{7}x - \tfrac{3}{7} = 0.$$

The next step is to obtain from this another equation whose roots are seven times as large. In accordance with the foregoing result this equation must be

$$x^2 + 4x - 21 = 0,$$

the roots of which are obviously -7 and $+3$. Hence the roots of the original equation are -1 and $+\tfrac{3}{7}$.

The same result is applied in Ex. XLIV, **No. 1**, where we have

$$ax^2 + bx + c = 0$$

$$x^2 + \frac{b}{a}x + \frac{c}{a} = 0.$$

The equation whose roots are $2a$ times the roots of this equation is

$$x^2 + 2bx + 4ac = 0$$

whence $$(x + b)^2 - (b^2 - 4ac) = 0$$

and $$x = -b \pm \sqrt{(b^2 - 4ac)}.$$

From this result we see at once that the roots of the original equation are $\{-b \pm \sqrt{(b^2 - 4ac)}\}/2a$.

§ 5. *Further Equations* (Ex. XLIV).—The examples of this exercise present more technical difficulty than those of the last. They include simple illustrations of equations with more than two roots. **Nos. 11-15** will be required in the discussion of ch. XXVII., B.

The graph of **No. 3** is shown in fig. 64. It is clear that $y = +4$ corresponds to no positive value of x. It is also obvious that it can correspond to no negative value, for if x is negative each of the two terms of the function is negative. The turning-point asked for may be read with approximate accuracy from the graph. It may be determined by calculation by the following argument :—

$$y = x + \frac{3}{x - 1}$$

$$x^2 - (y + 1)x + (y + 3) = 0$$

whence $x = \dfrac{y + 1 \pm \sqrt{\{(y - 1)^2 - 12\}}}{2}$.

Now this expression for x shows that there are no values of y which make $(y - 1)^2 - 12$ negative. The upper and lower limits of y are given, therefore, by the roots of the equation

$$(y - 1)^2 - 12 = 0,$$

that is, by $+ 1 \pm 2\sqrt{3}$. In the right-hand branch of the curve, therefore, the lowest value of y is $+ 4{\cdot}46$.

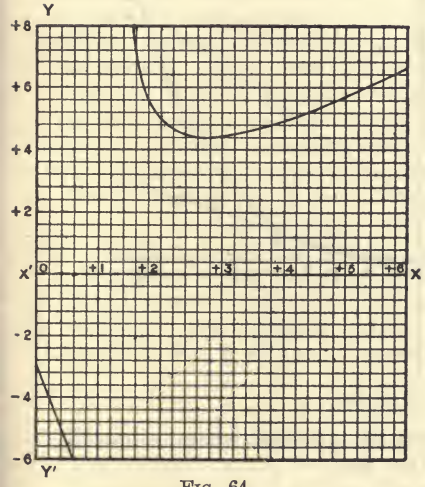


FIG. 64.

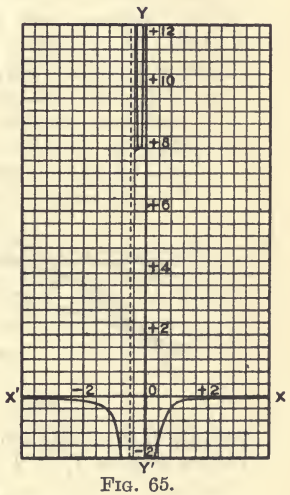


FIG. 65.

The graph of **No. 5** is fig. 65. There are evidently no values of x for which $y = 0$. The asymptotic values corresponding to $x = 0$ and $x = -0{\cdot}5$ should be noted with regard to both the upper and the lower branch of the curve.

In **No. 11** if the graph is moved 1 unit to the right we must substitute $x - 1$ for x; the corresponding function now becomes

$$y = (x - 1)^3 + 3(x - 1)^3 - 4(x - 1 - 1$$
$$= x^3 - 7x - 6$$

It is easy to see that by a similar substitution it is always possible to reduce the function

$$y = ax^3 + bx^2 + cx + d$$

to the form

$$y = ax^3 + c'x + d'$$

in which there is no term involving x^2. This property of a cubic function has great importance in the algebraic solution of equations of the third degree; it is also the basis of the graphic method of solution exemplified in **Nos. 12-15**. In fig. 66 the curve is the graph of $y = x^3$, the line AB that of

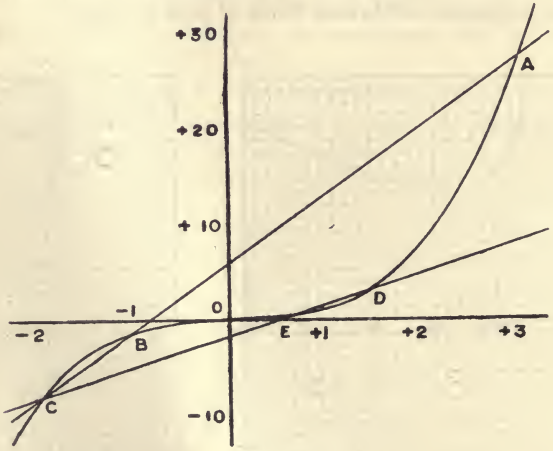


FIG. 66.

$y = 7x + 6$. The two graphs intersect where x has the values + 3, − 1, − 2. That is,

$$x^3 = 7x + 6$$
$$\text{or } x^3 - 7x - 6 = 0$$

when x has any one of those three values. But the roots of the original equation are lower by 1 than the roots of the derived equation; hence the roots of

$$x^3 + 3x^2 - 4x - 12 = 0$$

are + 2, − 2, and − 3.

In **No. 14** the graph when shifted 0·5 to the left corresponds to

$$y = x^3 - 3{\cdot}25x + 1{\cdot}5.$$

Thus we have to determine the intersection of the graphs of $y = x^3$ and $y = 3{\cdot}25x - 1{\cdot}5$. The latter is the line CED in fig. 66, and the points of intersection are where x has the values -2, $+0{\cdot}5$, and $+1{\cdot}5$. The roots of the original equation are, therefore, $-1{\cdot}5$, $+1$, and $+2$.

§ 6. *Inverse Functions* (Exs. XLV, XLVI; ch. xxvi., D). —The algebraic process for finding the roots of a quadratic equation may be looked at from a different point of view—namely, as a means of changing the subject of a formula in which one concrete variable is represented as connected with another by a parabolic relation. This use of the process is illustrated in division A of Ex. XLV. The natural sequel is to regard the formulæ obtained in this way as exemplifying a new set of functions, each of which is the "inverse" of the parabolic function represented by the formula from which it was derived. The formal study of the relations between direct and inverse parabolic functions and their respective graphs becomes, then, the subject of division B of the exercise. Both these matters are treated fully in ch. xxvi., D, but a few comments upon the examples may be of service.

The formula of **No. 3** can be thrown into the form

$$h = +72 - 0{\cdot}02(d - 4)^2$$

from which it is seen that the ball reaches its greatest height (6 feet) when 4 feet past the bowling crease. The answers to the other two questions are obtained by putting $d = 0$ and $d = +66$. The foregoing formula leads directly to the one required in **No. 4** with d as subject:—

$$(d - 4)^2 = (72 - h)/0{\cdot}02$$
$$d = +4 \pm \sqrt{(3600 - 50h)}.$$

The substitution of $h = 1{\cdot}45$ and $h = 12$ gives the answers to the questions set in No. 4.

The two methods of solving **No. 11** are those represented by the formulæ

$$S = 16\tfrac{81}{24} - \tfrac{3}{2}(n - \tfrac{41}{6})^2$$

and

$$n = \frac{+41 \pm \sqrt{(1681 - 24\,S)}}{6}.$$

Both show that S cannot rise above $1681/24$ or $70\frac{1}{24}$. The exact value of n corresponding to this value of S is $41/6$ or $6\frac{5}{6}$. Since, however, n must be integral it is clear that the greatest sum must actually correspond either to $n = 6$ or

$n = 7$. The second of these hypotheses makes $(n - \frac{41}{8})^2$ less than the former. We conclude that + 70, the value of S when $n = 7$, is its maximum. If the conclusion is correct the 8th and all subsequent terms must be negative so that their inclusion in the series would lower its value. This deduction is easily seen to be true.

The examples of **No. 16** are to be solved by the method explained at the end of ch. XXVI., D, § 2. Following the same rule in **No. 17** we have

$$y = ax + b$$
$$x = \frac{y - b}{a}$$

whence the inverse function must be

$$y = \frac{1}{a}x - \frac{b}{a}.$$

It will be found that the graphs of the original and inverse functions intersect where $x = y = b/(1 - a)$, that is, in a point whose co-ordinates are always equal. Such points can be found only on the line $y = x$.

Using the same method in **No. 18** we find that the function inverse to

$$y = \frac{a}{x + b} + c \text{ or } (x + b)(y - c) = a$$

is

$$y = \frac{a}{x - c} - b \text{ or } (x - c)(y + b) = a.$$

Here, without actually solving the equations, we can appeal to their "symmetry" as a proof that the values of x and y must be identical.

The argument of **No. 21** is of much importance in connexion with No. 22 and many subsequent examples.

In **No. 24** let a be any non-directed number less than $\frac{1}{2}$. Then when both $x = + a$ and when $x = - a$ the numerator of the function will be positive and the denominator negative. Thus the graph is below the x-axis both immediately to the right and immediately to the left of the y-axis. Also it is evident that it passes through the origin. It follows that the value of the function when $x = 0$, i.e. zero, must be an upper turning value. Next substitute for x the values $1 + a$ and

$1 - a$ in succession. The corresponding values of the function are easily found to be

$$3 + \frac{3a^2}{1 + 2a} \text{ and } 3 + \frac{3a^2}{1 - 2a}.$$

But the second term of each of these expressions is positive so long as $| a | < \frac{1}{2}$. Thus the graph is above the line $y = + 3$ on both sides of the line $x = + 1$. Also, when $a = 0$ (or $x = + 1$), $y = + 3$. We conclude, then, that when $x = + 1$ the value of the function $(+ 3)$ is a lower turning value.

No. 26 can be solved by the method exemplified on p. 241. Putting

$$y = \frac{x}{(x - 1)^2}$$

we have

$$x^2 - \left(2 + \frac{1}{y}\right)x + 1 = 0$$

whence

$$x = \frac{(2 + 1/y) \pm \sqrt{\{(4y + 1)/y^2\}}}{2}.$$

Since the number under the root sign must not be negative its lowest value is zero. But the fraction $(4y + 1)/y^2$ is clearly zero when $y = - \frac{1}{4}$, positive for all higher and negative for all lower values. Hence the function may have any value which is not below $- \frac{1}{4}$; that is, $- \frac{1}{4}$ is a lower turning value. Substitution shows that the corresponding value of x is $- 1$.

The problem may also be solved by the simpler argument illustrated in the case of No. 24. Substituting $- \frac{1}{4}$ for y in the original relation we find that $x = - 1$. For x substitute $- 1 + a$ in

$$y = \frac{x}{(x - 1)^2}$$

a being a small positive number; then the value of the function changes from $- \frac{1}{4}$ to

$$- \frac{1 - a}{(a - 2)^2}.$$

The numerator is now higher than before and the denominator lower; the value of the function is, therefore, raised. It follows that $- \frac{1}{4}$ is a lower turning value. Strictly speaking, neither of the foregoing arguments is complete, for they prove

only that the function has no values below $-\frac{1}{4}$. It is conceivable that the graph should run out in a single curved line to the point $(-1, -\frac{1}{4})$ and simply end there like an unfinished railroad. To prove that the point is a turning-point in the natural sense of the term, we must show that the graph approaches it along one course and leaves it along another. That this is the case follows from the consideration that to every value of y above $-\frac{1}{4}$ there correspond *two* values of x.

Nos. 27-30 are solved by similar arguments.

Ex. XLVI illustrates the application of the preceding methods to formulæ containing trigonometrical ratios. These examples may, perhaps, be claimed as evidence of the simplicity and economy which result from the assimilation of trigonometry with algebra. **Nos. 1-5** are instances of the "ambiguous case" in the solution of triangles. **No. 9** is a very simple proof of Euclid, III, 36. Since the constant term of a quadratic equation is the product of the roots we see that the rectangle contained by the two values of OP is $d^2 - r^2$. Inspection of a figure shows that this is equal to the square on the tangent. When O is within the circle d is less than r so that $d^2 - r^2$ becomes negative (**No. 10**). The interpretation of this result is that the two values of OP are the lengths of the lines drawn from O in opposite directions to meet the circle. That is, our result is now equivalent to Euclid, III, 35.

§ 7. *Wallis's Law* (Exs. XLVII, XLVIII, XLIX, ch. XXVII., A, B).—In the last two exercises of the present subdivision two important steps are taken towards the theory of the calculus. Ex. XLVII carries the doctrine of "integration" a stage forward from the simple introduction in Ex. XXX; Ex. XLVIII takes up the notion of an "approximation-formula," which we have had before us at intervals from Ex. IX onwards, and develops it into a more or less formal doctrine of "differentiation". Ch. XXVII. gives a full outline of the necessary exposition; we may, therefore, confine ourselves here to certain general considerations which the teacher should have before him.

The first point to note is that integration and differentiation are taught in these exercises as a "calculus of approximations". That is to say, our investigations, though giving results which may be regarded as true to any required degree of approximation, do not give, and must not be represented as giving, absolute results. This is a point of great import-

ance. There are two common views about the calculus, both erroneous and equally apt to cause confusion of ideas. The first is that the calculus is incapable of giving anything but approximate results although the error involved in them may be regarded as "infinitely" small. The other is that arguments which actually prove only an approximate result of this kind may be treated as if they had established an exact truth. The former error is pardonable, for it has its roots in the history of the subject. Until the mathematicians of the nineteenth century worked out the theory of "limits" it was not easy to see that, when properly stated, the arguments of the calculus do as a matter of fact give results which are as unequivocally exact as those obtained by multiplication or any other arithmetical process. For the latter error there is much less excuse, for it is due to a lack of precision in thinking which almost deserves to be called intellectual dishonesty. The appearance in an argument of either of the phrases "infinitely small" and "infinitely great" should always put the reader on his guard against an illegitimate deduction born of this vicious thinking; the occurrence of the word "ultimately" should make him actively suspicious. Nor let it be supposed that the logical lapses here in view are to be regarded as of trivial importance on the ground that "nobody seems a penny the worse" for them. On the contrary, one of the greatest hindrances to the beginner lies in the fact that arguments are so often presented to him as sufficient which he feels to be quite unconvincing. It is probable that the mature reader himself once felt the same discomfort when asked to assent to these specious reasonings. He has now ceased to be troubled by them chiefly because experience has given him so many opportunities of verifying the accuracy of results which he had originally to take partly and, perhaps, reluctantly on faith.

If the reader asks why the arguments which lead to exact conclusions are not taught in this book from the outset, the answer is that the doctrine of limits, though beautiful and entirely satisfactory, cannot profitably be presented to the beginner. It is better to reserve it, therefore, until familiarity with inferior methods of investigation has made him feel the need of something better. Moreover, it must be remembered that the results obtained by the lower type of reasoning are quite sufficient for all *practical* purposes. The demand for a

higher type commonly arises only when immediate practical needs are satisfied and the impulse which seeks its satisfaction in a completer theory begins to stir in the student's mind. The important thing is not to prejudice the success of this theoretical activity by encouraging or permitting an uncritical acceptance of an argument's claim to prove what in fact it does not and cannot prove. For these reasons care should be taken to keep the student awake to the exact significance of the conclusions he reaches.

In the second place the teacher will note that the traditional or Leibnizian notation is used in neither of the lessons of ch. XXVII. In the case of lesson A the reason is that the use of a technical notation would be unnecessary and would tend to distract the student's attention from the real business in hand. That business may be expressed as follows. The essence of the idea conveyed by the term "function" is that the value of one variable can be calculated from the value of another by the uniform application of a definite rule expressible in algebraic symbolism. In addition to this idea our studies have made us familiar with another notion, namely, of a function whose algebraic form can be calculated from the form of another *function* by the uniform application of a definite rule. As a simple example let y be a variable whose value is calculable from the value of x by the function $2x - 3$, and let it be required to find a function of x by which the value of the variable y^2 may be calculated. Then we know that this function can be derived from the function $2x - 3$ by the application of a perfectly definite and universal rule: it will, in fact, be $4x^2 - 12x + 9$. It is easy, if we wish to do so, to give a graphic expression to the connexion between these two functions. If any given value of the former is represented by a straight line of the proper length the corresponding value of the latter will be represented by the area of the square drawn upon it. Now there are many important problems in which, if the successive values of the first of two functions are represented by the ordinates of a certain curve, the corresponding values of the second function will be represented by the area under the curve from the y-axis up to the successive ordinates. The question then arises whether it is possible in these cases, as in the former example, to lay down a definite rule from which, given the form of the "ordinate-function," the form of the "area-function" can be at once

determined. The contribution of ch. XXVII. to the solution of this question is the proof that in certain cases, at any rate, such a rule can be given. If the "ordinate-function" is of the form kx, or kx^2, or kx^3, then the "area-function" has the form $\frac{1}{2}kx^2$, or $\frac{1}{3}kx^3$, or $\frac{1}{4}kx^4$ as the case may be. Or, in general, if the original function is of the form

$$y = a + bx + cx^2 + dx^3$$

then the second function is of the form

$$A = ax + \tfrac{1}{2}bx^2 + \tfrac{1}{3}cx^3 + \tfrac{1}{4}dx^4.$$

This important rule we call, in memory of its discoverer, "Wallis's Law". In order that its true significance as a rule by which in certain arguments we may pass from a function of one form to a function of another form may be kept clearly before the student, it is well to avoid a technical notation whose original meaning was very different from this. A graphic presentation is by far the best means of keeping in mind the conditions under which the rule is applicable.

In ch. XXVII., B, and Ex. XLVIII, the case is different. Here the argument requires a definite notation to symbolize a new idea—namely, the idea expressed verbally by the term "differential" with its carefully defined connotation of approximate equality. The Leibnizian notation dy/dx is not used for this purpose because, when correctly understood, it does *not* symbolize an approximation, but expresses an *exact* equality. It is reserved, therefore, for Section VIII where it can be introduced without danger of confusion. The same remark applies to the Leibnizian symbol for integration. We may add that for the same reason the terms "integration" and "differentiation" are not given to the pupil until he reaches Section VIII. They are employed in these observations merely because they are familiar to the reader. It is desirable that the student should be taught to use them only in connexion with results based upon the theory of limits.

The teacher will recognize the application of these remarks in the lessons of ch. XXVII. In lesson A Wallis's favourite argument is used to show that if $(m + 1)$ columns of equal width whose heights are successively proportional to the numbers 0^2, 1^2, 2^2, 3^2, . . . m^2 are laid upon the rectangle AM (fig. 76) they will cover a fraction of its area which is given by the expression

$$\frac{1}{3} + \frac{1}{6m}.$$

It is certain that, as m increases, this fraction becomes constantly nearer to 1/3, and that by taking m sufficiently large the approximation may be made as close as we please. It also seems obvious that the area covered by the columns approximates in much the same way to the area under the semi-parabola OP. Thus having *proved* that the area under the curve is, apparently to an unlimited degree of closeness, one-third of the area of the rectangle AM, we are almost forced to believe that the former is exactly one-third of the latter. At any rate if the fraction is not truly one-third there is practically no likelihood that the discrepancy could be detected by measurement. Still the fact remains that what we have proved is only an approximation. However numerous the columns are made they will remain a set of columns and their tops can never become a parabolic curve. It remains an assumption that, because their united area *approaches* one-third of AM, the area under the parabola is *exactly* one-third of that area. Thus while we may legitimately emphasize the practical certainty of our conclusion no attempt should be made to disguise its theoretical imperfection.

Similarly, in ch. XXVII., B, and Ex. XLVIII, no attempt should be made to slur over the fact that the "differential formula" is always, from the logical standpoint, merely an approximation-formula, though it is quite proper to emphasize the other fact that no limit can be set to the closeness of the approximation which it represents. In the interests of lucidity and exact thinking it is equally important that careful attention should also be given to the points brought out in the Notes before Nos. 12, 13, 15 and 17.

It has already been remarked (p. 57) that in the investigation of his law Wallis was led to the invention of negative and fractional indices. In the same place it was said that, on the whole, it is better to introduce the theory of fractional indices in connexion with logarithms. The class which follows either of the schemes set out on p. 50 will probably reach the theory of indices in Section III before it reaches Ex. XLVII. Nevertheless, the index method of representing a square root has been introduced in the argument of ch. XXVII., A, upon Wallis's lines and without reference to the more formal treatment of ch. XXXIV. This procedure leaves the teacher free to choose an order of treatment different from either of those given on p. 50, and also enables the student to

become acquainted with an argument of great intrinsic interest and historical importance.

In Ex. XLVII, **No. 1** (vi), the area-function is, by Wallis's Law,

$$A = 3{\cdot}4\ (-x)^{\frac{3}{2}}$$

the negative sign implying that the whole of the curve is on the left of the y-axis. In **No. 5** the area-function is $A = x^{\frac{3}{2}}$. Hence the areas under the curve up to the ordinates $x = +49$ and $x = +4$ are respectively

$$(+49)^{\frac{3}{2}} = +343 \text{ and } (+4)^{\frac{3}{2}} = +8.$$

Thus the area between these ordinates is $343 - 8 = 335$.

In solving **No. 6** the easiest method is to apply directly the theorem that the area within the curve is two-thirds of that of the rectangle with the same height and base. Thus the height of A above the x-axis is 25·6 and the distance between the points where the curve cuts the x-axis is 32; hence the required area is

$$\tfrac{2}{3} \text{ of } 25{\cdot}45 \times 32 = 542{\cdot}9.$$

In **No. 7** it will, of course, be necessary to find two parabolic areas in this way and to take their difference.

Nos. 9-11 are important as leading to the conclusion that Wallis's Law can be used to find the area-function when the ordinate-function is a complex of the form

$$y = a + bx + cx^2 + \ldots .$$

In **No. 9** we find that

$$pq = 0{\cdot}04x^2$$

whence we deduce that the area between AB and AC (*Exercises*, fig. 44) may be regarded as the area under the parabolic curve $y = 0{\cdot}04x^2$ after every ordinate has been shifted vertically upwards—just as the dark columns in fig. 55 (p. 206) have been shifted vertically to produce fig. 56. It follows that the area-function of the surface between AB and AC is the same as that of the space under $y = 0{\cdot}04x^2$, namely,

$$A = \tfrac{1}{75}x^3$$

and that the total area is one-third of that of a rectangle whose sides are equal to AD and BC. To answer the first part of **No. 10** we have merely to add the area of the triangle ADC. But the ordinate-function for points on AC is easily seen to be $y = 0{\cdot}8x$; hence the ordinate-function for points on AB (**No. 11**) must be

$$y = 0{\cdot}8x + 0{\cdot}04x^2$$

In view of the preceding argument the area-function must be

$$A = \tfrac{2}{5}x^2 + \tfrac{1}{75}x^3.$$

Substituting $x = 20$ we obtain the same result as before for the total area under the curve AB.

In **No. 21** the successive fractions can be arranged in a sequence in which the denominators show a constant difference of 4 :—

$$\tfrac{2}{4}, \tfrac{3}{8}, \tfrac{4}{12}, \tfrac{5}{16}, \tfrac{6}{20}, \ldots .$$

It is obvious that they are all of the form

$$\frac{m+1}{4m} = \frac{1}{4} + \frac{1}{4m}.$$

Assuming in **No. 22** that this law will hold good for all values of m, it follows that the area under the curve will be one-quarter of that of the underlying rectangle. In **No. 30** this assumption is to be justified by the method of recurrence already illustrated in ch. XXVII., A, 3. The proof is as follows.

Assume that

$$\frac{0^3 + 1^3 + 2^3 + \ldots + p^3}{(p+1)p^3} = \frac{p+1}{4p}$$

that is, that

$$0^3 + 1^3 + 2^3 + \ldots + p^3 = \tfrac{1}{4}(p+1)^2p^2.$$

Add the term $(p+1)^3$ to each side. Thus we have

$$0^3 + 1^3 + 2^3 + \ldots + (p+1)^3 = \tfrac{1}{4}(p+1)^2p^2 + (p+1)^3$$
$$= \tfrac{1}{4}(p+2)^2(p+1)^2$$

as in Ex. XXXVIII, No. 10. It follows that

$$\frac{0^3 + 1^3 + 2^3 + \ldots + p^3 + (p+1)^3}{(p+2)(p+1)^3} = \frac{1}{4} + \frac{1}{4(p+1)}.$$

We conclude that if the result in question holds good when $m = p$ it also holds good when $m = p + 1$. But it is known by trial to hold good when $m = 1, 2, 3 \ldots$. Therefore it holds good universally.

Practically all the difficulties likely to be met with in Ex. XLVIII have been anticipated in the exposition of ch. XXVII., B.

In Ex. XLIX, fig. 67 is the graph of **Nos. 2 and 8.** From the formula

$$y = x^3 - 3x + 2$$

we deduce that

$$\frac{\delta y}{\delta x} = 3x^2 - 3 \quad . \quad . \quad . \quad A$$

and that

$$\frac{\delta^2 y}{\delta x^2} = 6x \quad . \quad . \quad . \quad . \quad \text{B}$$

The gradients at the points specified in **No. 1** are found by substituting the given values of x in A. For example, when $x = -1$, $\delta y/\delta x = 0$; that is, the point is a turning point. If x is a little below -1, $\delta y/\delta x$ is positive, if a little above, negative. It is clear, therefore, that this turning point is a maximum. This fact is also proved (**No. 8**) by the circumstance that $\delta^2 y/\delta x^2$ is negative when $x = -1$. At $x = 0$, $\delta y/\delta x = -3$, and $\delta^2 y/\delta x^2 = 0$. If x is taken a little below zero $\delta^2 y/\delta x^2$ is negative, if a little above, positive. It follows (**No. 8**) that the intersection of the graph with the y-axis is a point of inflexion.

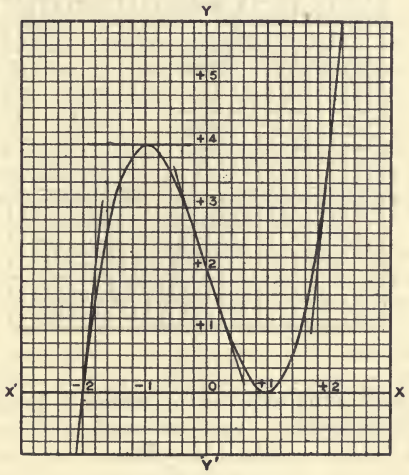


FIG. 67.

Fig. 68 is the graph of **Nos. 4 and 9**. Here we have

$$\frac{\delta y}{\delta x} = 6x^2 + 6x - 36 \quad . \quad . \quad . \quad \text{A}$$

$$\frac{\delta^2 y}{\delta x^2} = 12x + 6 \quad . \quad . \quad . \quad . \quad \text{B}$$

As before, the gradients at the points specified in **No. 3** are

to be found by substituting the given values of x in A, while the character of the curvature at each point may be examined either by considering how the gradient varies to right and left of it or by substituting for x in B. For $x = -0{\cdot}5$ $\delta^2y/\delta x^2 = 0$; for a value a little below this, $\delta^2y/\delta x^2$ is negative, for one a little above, positive. Thus the point is a point of inflexion; the curve is below the tangent on the left of it and above the tangent on the right.

In **No. 22** let a be the side of a square end and l the length

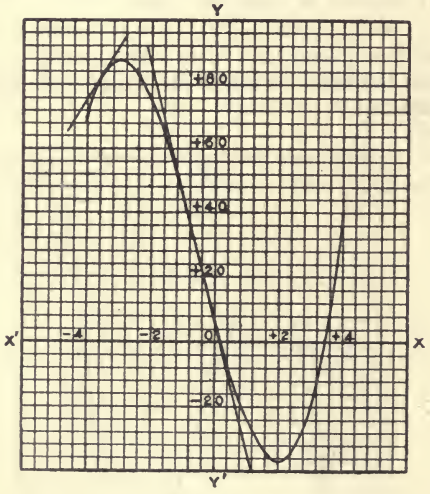


FIG. 68.

of the box, both being measured in inches. Also let V be the volume in cubic inches. Then we have

$$4a + l = 72$$

$$V = a^2l$$

$$= 72a^2 - 4a^3$$

$$\frac{\delta V}{\delta a} = 144a - 12a^2 \quad . \quad . \quad . \quad A$$

$$\frac{\delta^2 V}{\delta a^2} = 144 - 24a \quad . \quad . \quad . \quad B$$

From A we have that $\delta V/\delta a = 0$ when $a = 0$ or $a = 12$. From B we see that $a = 0$ makes $\delta^2V/\delta a^2$ positive, while

$a = 12$ makes it negative. Hence the latter value implies a maximum. The corresponding value of l is 24 inches. Thus the largest box is 2 feet long with an end 1 foot square.

§ 8. *Directed Trigonometrical Ratios* (Exs. XL, XLI; ch. xxv., B).—We now turn to the consideration of the exercises and lessons in which the idea of directed numbers is applied to the trigonometrical ratios. The case of the tangent has been dealt with already (§ 2). The principle followed there is applied again in the case of the sine and cosine. It is shown that certain formulæ would become simplified and (so to speak) condensed if it could be supposed that angles between 90° and 180° have, like acute angles, sines and cosines. These new sines and cosines are then defined in such a way that the desired simplifications are secured. Thus we decide that the sine of an obtuse angle shall be considered identical with the sine of its supplement, and the cosine equal to the cosine of the supplement but opposite in sign. It is important to observe that no proof is or can be offered of these statements; they are simply conventions adopted in order to bring different cases of a rule (such as the rule for finding the area of a triangle) under a single formula. The examples of Ex. XL give practice in the application of these new sines and cosines and also in the two important general properties which are proved in the course of the lesson.

From the theoretical standpoint the importance of Ex. XLI is that it carries still farther the principle just explained. It is shown that the work of the surveyor is much simplified by assigning sines and cosines to every possible angle from 0° to 360°. Once more the new sines and cosines must be defined in such a way that they actually perform the service which is required of them and for which they are called into being. The guiding idea this time is that it shall be possible to find correctly the northing or southing and the easting or westing (ch. xiii., B) which correspond to any given vector by multiplying its length, in the first place by the cosine, and in the second case by the sine, of its bearing, bearings being measured continuously round from the north through the east. The values of the sines and cosines to be allotted to the various angles are to be determined in such a way that this convenient result follows. The exercises close with a few examples (forming division C) to introduce the idea of a negative angle which may have any value up to 360° and

may also be thought of as possessing a sine, cosine, and tangent.

This extension of the field of the sine, cosine, and tangent marks the limit reached in Part I. Any further extension should be based, as these extensions have been, upon the requirements of some practical problem. Such problems do not arise until, in Part II, Section VII, we meet them in connexion with the study of harmonic motion and waves.

Ex. XLI is also intended to illustrate with a certain amount of actuality the methods used by the surveyor in mapping out a country. It is necessary that the work in some examples should be divided among the members of the class or it may become undesirably burdensome. If possible, the examples should be supplemented by simple field work carried out as nearly as may be in accordance with the actual methods of the surveyor. Inexpensive prismatic compasses and theodolites are now to be obtained from several makers of scientific apparatus. In boys' schools there should be little difficulty in supplying efficient instruments out of the resources of the manual training department.

§ 9. *The Calculation of* π. *Trigonometrical Tables* (Ex. L; ch. XXVIII., A, B).—Even an elementary course must be considered incomplete if it includes no discussion of the methods by which trigonometrical tables may be calculated. This highly interesting and important subject is generally omitted from elementary treatises because it is thought to presuppose a knowledge of the expansions for the sine and cosine of a given angle. Yet it is probable that all the tables actually in use in our schools are lineally descended from tables calculated long before the conception of these expansions had entered into any mathematician's head. Archimedes obtained his famous approximations for π by considering the perimeters of inscribed and circumscribed polygons. Hipparchus (*c.* 150 B.C.) probably, and Ptolemy (*c.* A.D. 100) certainly, calculated the lengths of the chords which correspond to the various angles of the quadrant. These were used in trigonometrical calculations right down to the later middle ages when the use of semi-chords or sines filtered into Europe from the Arabs. Tangents were invented by the celebrated Regiomontanus about 1450; secants followed about a century later, but the credit of their invention is disputed.

The first modern method of calculating π is that of Ludolph

van Ceulen ; our own John Wallis followed in 1655 with a totally different method which gave the value as an infinite product. Van Ceulen's method and the method of calculating trigonometrical tables which goes with it are illustrated in ch. XXXVIII. and Ex. L. The exposition of the chapter is too full and the examples too straightforward to need comment.

Archimedes' famous evaluation of π is given in full in Sir T. L. Heath's translation of his collected works. Much interesting information about the general history of the trigonometrical ratios is given in the introduction to Hutton's *Mathematical Tables* (1785). Wallis's account of his discovery of the product-approximation for π is given in his *Arithmetica Infinitorum* (1655). The present writer has given a summary of it in the *Mathematical Gazette* for Dec. 1910 and Jan. 1911. Van Ceulen's book, as far as he knows, remains inaccessible except to readers of the original Dutch or of Snell's Latin translation of 1619. A few of the huge numbers with which it teems are given in ch. XXXVIII. to illustrate the immense labours which the founders of modern mathematics were sometimes ready and even glad to face. His longest value for π—the "Ludolphian number" as it was called—is said to have been inscribed upon his tomb. He was modestly proud of his achievements, and records with an obvious swelling of the soul the precise date upon which his method occurred to him : "Anno reparatæ salutis 1586 mense septembri istam ad laterum polygonorum circulo adscriptorum investigationem viam inveni"; and relates that he immediately proposed to himself the tremendous task of calculating the perimeters of the inscribed and circumscribed polygons of 167,772,160 sides !

CHAPTER XXV.

LINEAR FUNCTIONS. EXTENDED USE OF SINE, COSINE AND TANGENT.

A. *Linear Functions* (ch. XXIV., § 2; Ex. XXXIX).

§ 1. *The Relation* $y = a + bx$ *always Linear.*—It is obvious that the relation $y = kx$ describes a straight line through the origin even when k, as well as x and y, is directed. For example, $y = +0{\cdot}9x$ describes the line AA′ (fig. 69) and $y = -0{\cdot}9x$ the line BB′. The proof is essentially that of ch. XII., A, § 2. In $y = +0{\cdot}9x$ let the values of x rise by equal steps from zero to OM. Then the value of y rises by equal steps, 0·9 times those of x, to PM. On the other hand if the value of x falls by equal steps from zero to OM′, y falls by equal steps, 0·9 times those of x, to P′M′. Thus the graph is a continuous straight line lying in what may be conveniently called the *first* and *third* quadrants of the graph-paper. The argument may be repeated with regard to $y = -0{\cdot}9x$ with the difference that upward steps in x imply downward steps in y and *vice versa*, so that the corresponding straight line, BB′, lies in the second and fourth quadrants. Similar results could obviously be obtained for all values of k, positive or negative.

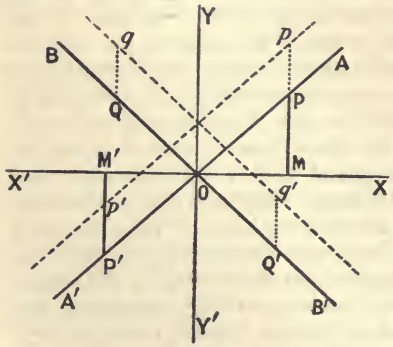


FIG. 69.

Consider next the relations

$$y = + 0{\cdot}9x + 5 \text{ and } y = - 0{\cdot}9x + 5.$$

It is clear that the effect on the graph of the addition of + 5 to + $0{\cdot}9x$ in one case and − $0{\cdot}9x$ in the other is to raise each ordinate through 5 units of the vertical scale. Thus P and P′ rise to p and p', Q and Q′ to q and q'. We may say, then, that the new relations describe the old lines raised, parallel to themselves, through 5 units. Similarly $y = + 3x - 7$ describes the line through the origin, $y = + 3x$, after it has been lowered 7 units.

We conclude that a relation of the form $y = px + q$ is a linear relation for all values of p and q, and that it describes the straight line through the origin, $y = px$, moved through q units—upwards if q is positive, downwards if q is negative.

§ 2. *Tangents as Directed Numbers.*—AA′ and BB′ (fig. 69) may be described as lines in which the ordinates bear a constant ratio to the abscissæ—the ratio being in the former case + 0·9, in the latter, − 0·9. This description is an obvious extension to directed variables of the notion of direct proportion (ch. XII., A). But the lines, as lines, can be described in another manner. The table on *Exercises*, p. 107, gives $0{\cdot}9 = \tan 42°$. AA′ and BB′ are, therefore, lines inclined at 42° to the x-axis, the difference between them being only in the direction of the inclination. Both lines might be supposed to have started from the position XX′ and to have reached their present position by rotating about O through 42°, AA′ in the anticlockwise and BB′ in the clockwise direction. Thus, speaking generally, the relation $y = kx$ would describe a line through the origin having a certain inclination to the x-axis. The inclination would be the angle whose tangent is numerically equal to k, and it would be anticlockwise if k were positive and clockwise if k were negative.

But there is a disadvantage in this way of describing the line $y = kx$. Suppose a line to start from XX′ and to be rotated anticlockwise about O until it coincides with YY′. Then k in $y = kx$ will assume successively all possible positive values from zero upwards. According to the last paragraph, if the line is now to assume the various negative inclinations in due order it must go back to XX′ and begin to rotate again in the opposite direction. Now it seems much more natural to suppose the line to reach the position

BB′ by a continuation of the process which brought it through AA′ to YY′—that is, by a continuous anticlockwise rotation. To secure this simpler way of looking at the matter all we need do is (*a*) to measure the "inclination" of the line by supposing in every case an anticlockwise rotation from XX′; and (*b*) to extend the idea of a tangent to include negative and positive values. Thus the tangents corresponding to rotations of 100°, 110°, 120°, etc., must be taken to be − 5·67, − 2·75, − 1·73, etc.; that is to say, the tangents of 80°, 70°, 60°, etc., with a negative sign prefixed. In general, if α is an angle between 90° and 180° we must assume (or rather *define*) that

$$\tan \alpha = - \tan (180° - \alpha).$$

Upon this understanding it becomes easy to lay down a single rule for describing the line which corresponds to any given linear relation. For example, $y = + 0{\cdot}7x - 3{\cdot}7$ describes a straight line, inclined 35° to the x-axis (shortly, "of inclination 35°"), which has been lowered from the standard position through 3·7 units; $y = - 0{\cdot}7x + 12{\cdot}8$ describes a line, inclined 145° (= 180° − 35°) to the x-axis, which has been raised through 12·8 units.

§ 3. *Linear Functions.*—We have constantly spoken of a *relation* between two variables, meaning that when one variable changes the other variable also changes in some definite way. This kind of connexion between variables is often described by saying that one is a *function* of the other. The idea of a function as it is used in mathematics is perfectly simple; only the name is alarming. Thus the length of a rod is said to be a function of its temperature because if you change the temperature the length will change. The rent a man pays for his house is (or should be) a function of two variables—his income, and the size of his family. If he has a larger or smaller income or a larger or smaller family he will normally pay a larger or smaller rent. According to uncompromising temperance advocates the amount of crime in England is almost entirely a function of the amount of alcohol consumed. As a last example we may take the celebrated statement that the quantity of red clover on a farm is a function of the number of its cats. For red clover depends for fertilization upon humble-bees which are preyed upon by field-mice which in their turn are the prey of the cats!

It is obvious that one variable may be a function of any number of variables. For example, the national income obviously depends upon a vast number of changing factors, each of which by rising or falling in amount affects the total in some way. For the present, however, we shall consider only functions of a single variable, and we shall denote that variable by x. By speaking of a function of x we shall mean, then, some expression whose value depends on the value of x. Thus

$3x + 7$, $\sqrt{9 \cdot 2x - 3 \cdot 4}$, $x^3 - 2x + 4$, $1/(4 - 3x + 2x^2)$

are all functions of x. The value of the function for a given value of x will always be denoted by y. Thus the expression $y = \sqrt{9 \cdot 2x - 3 \cdot 4}$ may, as we please, be regarded either as stating that a certain relation exists between the values of x and y, or as stating that y is the value of a particular function of x, namely $\sqrt{9 \cdot 2x - 3 \cdot 4}$. The graph of a given relation between x and y may also be described as the graph of the corresponding function of x. Thus since the graph corresponding to any relation of the form $y = px + q$ is a straight line we may say that any function of x which can be reduced to the form $px + q$ is a *linear function*.

[Ex. XXXIX may now be worked.]

B. Extension of Meaning of Sine and Cosine (ch. XXIV., § 8; Exs. XL and XLI).

§ 1. *The Problem.*—In ch. XIII., B, the sine and cosine were studied as factors by which the sides of a right-angled triangle can be calculated, given the hypothenuse. Since the angles of such a triangle cannot be greater than 90°, only angles between 0° and 90° were considered as having sines and cosines. The advantage of assigning tangents to angles between 90° and 180° now suggests the inquiry whether the range of sines and cosines is capable of a similar useful extension.

§ 2. *Sines of Angles between* 90° and 180°.—Let us take the sine first. Fig. 70 shows that, when the angles A and B are both acute, we have:—

$$\text{area of triangle} = \tfrac{1}{2}cp = \tfrac{1}{2}ca \sin\beta = \tfrac{1}{2}cb \sin\alpha.$$

But, when the angle A is obtuse (fig. 71), while the equivalence

$$\text{area} = \tfrac{1}{2}ca \sin\beta$$

still remains, the second formula becomes

$$\text{area} = \tfrac{1}{2}bc \sin(180° - \alpha).$$

That is to say, we cannot express the area of a triangle in terms of two sides and the included angle without first inquiring whether the angle is acute or obtuse; the form of the expression being different in the two cases.

This inconvenience could be removed by agreeing that angles between 90° and 180° shall be considered to have sines, the rule being that $\sin \alpha = \sin (180° - \alpha)$—or, in words, that the sine of an angle shall be reckoned the same as the sine of its *supplement*. Upon this understanding the area of a triangle can always be expressed as $= \frac{1}{2} bc \sin \alpha$ whether α is acute or obtuse.

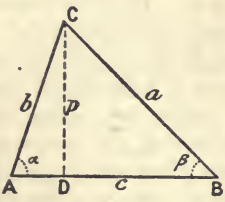


FIG. 70.

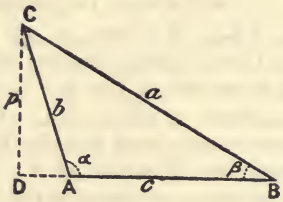


FIG. 71.

§ 3. *Cosines of Angles between* 90° *and* 180°.—The same figures suggest a similar extension of the cosine. In fig. 70 we have

$$c = b \cos \alpha + a \cos \beta$$

but in fig. 71 the formula must be written:—

$$c = a \cos \beta - b \cos (180° - \alpha).$$

As before, the inconvenience of two rules—one for an acute angled and another for an obtuse angled triangle—may be avoided by supposing that every angle between 90° and 180° has its own cosine as well as its own tangent and sine. But the rule takes this time the form

$$\cos \alpha = - \cos (180° - \alpha)$$

or, in words, that the cosine of an angle is to be reckoned the negative of the cosine of its supplement.

§ 4. *Harmony of these Extensions.*—For angles between 0° and 90° $\tan \alpha = \sin \alpha / \cos \alpha$. Does this result still hold good when α is between 90° and 180°? In such a case we have

$$\tan \alpha = - \tan (180° - \alpha)$$
$$\sin \alpha = \sin (180° - \alpha)$$
$$\cos \alpha = - \cos (180° - \alpha)$$

so that $\sin \alpha / \cos \alpha = - \tan (180° - \alpha) = \tan \alpha$ as before.

It is easy to redefine the sine, cosine and tangent of any angle between 0° and 180° in such a way as to secure all the results of ch. XIII. and the present chapter. Let the radius of the semicircle of fig. 72 be taken as unity and let the angle α be traced by the radius OP starting from the position OA. When it stops in any position such as OP_1 or OP_2 drop a perpendicular P_1N_1 or P_2N_2. Then PN is to be defined as sin α, ON as cos α and PN/ON as tan α. It is clear that for $\alpha < 90°$ (as AOP_1) sin α, cos α and tan α are all positive, while for $\alpha > 90°$ sin α is positive, cos α and tan α negative. Moreover it is obvious from the symmetry of the semicircle about OC that the sine, cosine and tangent of any angle are numerically the same as those of the supplement of the angle.

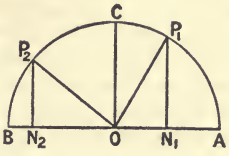


FIG. 72.

§ 5. *Two Important Theorems.*—Two important properties of the triangle can now be stated very concisely :—

(*a*) Whether a triangle be acute angled or obtuse angled we have :—

$$\text{area} = \tfrac{1}{2}bc \sin \alpha = \tfrac{1}{2}ca \sin \beta = \tfrac{1}{2}ab \sin \gamma$$

Hence (multiplying by $2/abc$) :—

$$\sin \alpha/a = \sin \beta/b = \sin \gamma/c$$

or : each side is directly proportional to the sine of the opposite angle.

(*b*) Whether the triangle be acute angled or obtuse angled [1] we have (figs. 70 and 71) :—

$$\begin{aligned} a^2 &= (b \sin \alpha)^2 + (c - b \cos \alpha)^2 \\ &= b^2 \sin^2 \alpha + c^2 - 2bc \cos \alpha + b^2 \cos^2 \alpha \\ &= b^2 (\sin^2 \alpha + \cos^2 \alpha) + c^2 - 2bc \cos \alpha \\ &= b^2 + c^2 - 2bc \cos \alpha \end{aligned}$$

The former theorem makes it possible to solve many problems more conveniently than could otherwise be done; the latter enables us for the first time to calculate any angle of a triangle when the lengths of its sides are known.

[Exs. XL and XLI can now be taken.]

[1] In fig. 71 $p = b \sin (180° - \alpha) = b \sin \alpha$; $AD = b \cos (180° - \alpha) = b \times (- \cos \alpha)$.

CHAPTER XXVI.

HYPERBOLIC AND PARABOLIC FUNCTIONS.

A. Hyperbolic Functions (ch. XXIV., § 3; Ex. XLII, A).

§ 1. *The Rectangular Hyperbola.*—In ch. XXV., A, we saw how the use of directed numbers leads to an important extension of the notions of direct proportion and of the linear relation. We now inquire into the effect of admitting these numbers into the relation called inverse proportion.

For this purpose take the relation $xy = 24$ and plot the values of y corresponding to negative as well as positive values of x. Fig. 73 shows the result. In addition to the curve AVB, familiar from ch. XVI., A, we have, in the opposite quadrant, the curve A′V′B′. The second curve is a precise reproduction of the first, except in position; for every point on AVB, such as $(+ 6, + 4)$, is balanced by a point in A′V′B′, such as $(- 6, - 4)$. If a mirror were fixed along MM, at right angles to the axis VV′ and facing AVB, the image of AVB would be A′V′B′.

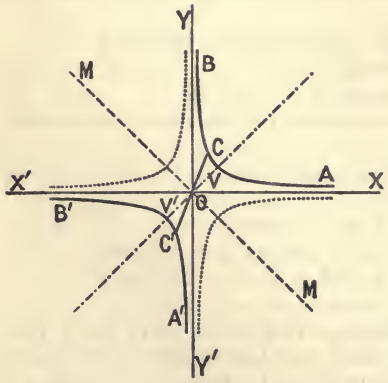


FIG. 73.

Although AVB and A′V′B′ are separated and lie in distinct quadrants yet they must be regarded as simply two branches

of one and the same curve, just as OA and OA′ in fig. 69 are parts of one line. Suppose a point, Q, to approach O along X′O from an endless distance beyond X′. Then the point, P, defined by the relation $y = + 24/x$, will move along B′V′A′. As the distance QO diminishes the length of the ordinate increases without end. That is, it is never so long that it does not become still longer as Q comes nearer still to O. As Q passes from one side of O to the other the story of the tangent is repeated (ch. XIII., A). We cannot say that the curve has a point *on* the y-axis, for it is senseless to speak of the quotient of $24 \div 0$. Yet no interval can be specified to the left and right of O so short that there are no corresponding positions of P still nearer to the y-axis—on the left at an endless depth along the y-scale, on the right at an endless height. As Q proceeds along OX, P simply repeats in reverse order along BVA its previous adventures along B′V′A′.

The one curve with its two branches is called a *rectangular hyperbola.* V and V′ are the *vertices*, the line VV′ $(= 2\sqrt{2}.k)$ is the *axis*, XX′ and YY′ the *asymptotes* (ch. XVI., A, § 2), O (because it bisects all chords such as CC′) the *centre.*

§ 2. *Movements of the Hyperbola.*—If the hyperbola is raised (say) through 7 units the corresponding relation must become $y = 24/x + 7$, or $y - 7 = 24/x$. If it is lowered 7 units we have $y = 24/x - 7$, or $y + 7 = 24/x$. Moving the figure to the left will not (as in Ex. XL, No. 12) produce the same effect as raising it, but it will affect the algebraic description of the curve in the same way as in the case of the straight line. If (for example) the curve is moved 3 units to the left the distance of a point from the vertical asymptote is no longer x; it becomes $x + 3$. Similarly a movement of 3 units to the right changes the distance from the asymptote from x into $x - 3$. Thus if the figure is moved 3 units to the right and then 7 downwards the relation becomes:—

$$y + 7 = 24/(x - 3) \quad \text{or} \quad y = 24/(x - 3) - 7.$$

§ 3. *Negative Values of k.*—If we had plotted $y = -24/x$ instead of $y = + 24/x$ the result would evidently have been to give the dotted curve of fig. 73. Two movements (for example) which carry the centre to the point $(-6, +5)$ change the corresponding relation to:—

$$y - 5 = -24/(x + 6) \text{ or } y = -24/(x + 6) + 5.$$

§ 4. *Hyperbolic Functions.*—Just as $ax + b$ is termed a

linear function of x because the graph of $y = ax + b$ is a straight line, so we may term any function that can be thrown into the form $k/(x + a) + b$ a *hyperbolic function of* x because the graph of $y = k/(x + a) + b$ is a rectangular hyperbola. This function is by no means unusual in physics where it occurs in cases that admit of negative and positive values of x.[1]

[Ex. XLII, A, may now be worked.]

B. Parabolic Functions (ch. XXIV., § 3; Ex. XLII, B).

§ 1. *The Parabola.*—Let us now explore the wider range included in the relation $y = kx^2$ when x, y and k may be directed numbers. The hard line of fig. 74 is the graph of $y = +\frac{1}{4}x^2$. Like the extended graph of $y = k/x$ it lies in two quadrants, but they are adjacent instead of opposite quadrants. The reason is obvious. Whether $x = p$ or $x = -p$, in each case $\frac{1}{4}x^2 = \frac{1}{4}p^2$. Thus OP′ is the image of OP in a mirror set up along OY. The curve extends upwards and outwards both ways without limit but no part of it lies below the x-axis. It is called a *parabola*. V (here identical with the origin O) is the *vertex* or head; VA (here identical with the y-axis) is the *axis* of the parabola about which the curve balances; TVT′ (here identical with the x-axis) is clearly a *tangent at the vertex*.

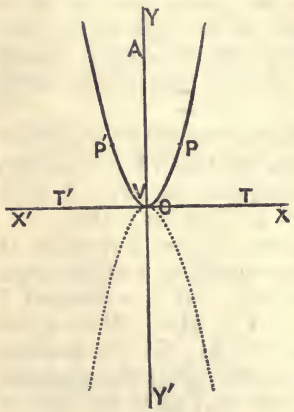


FIG. 74.

If we plot $y = -\frac{1}{4}x^2$ instead of $y = +\frac{1}{4}x^2$ we obtain the dotted curve, identical with the former except that it is "head up" instead of "head down".

[1] For example, the familiar formula $\frac{1}{v} + \frac{1}{u} = \frac{1}{f}$ which gives the distances of object and image from a lens. Substituting x for u and y for v we have $1/y + 1/x = 1/f$, becoming $(x - f)(y - f) = f^2$, or $y = f^2/(x - f) + f$.

§ 2. *Movements of the Parabola.*—Like the hyperbola the parabola can be supposed moved so that the axis and tangent at the vertex remain unchanged in direction although altered in position. Thus $y - 7 = \frac{1}{4}(x + 5)^2$ or $y = \frac{1}{4}(x + 5)^2 + 7$ describes the parabola $y = + \frac{1}{4}x^2$ moved 5 units to the left and 7 units upwards. So $y + 8 = -\frac{1}{4}(x - 10)^2$ or $y = -\frac{1}{4}(x - 10)^2 - 8$ describes $y = -\frac{1}{4}x^2$ moved 10 units to the right and 8 downwards. The shortest way to express the movements is to say that in the first case the vertex is now at $(-5, +7)$ and in the second case at $(+10, -8)$.

§ 3. *Parabolic Functions. Turning Values.*—In general, if we start with a parabola $y = ax^2$ and move it horizontally through b units, and vertically through c units (a, b, and c being all directed), it corresponds in its final position to $y = a(x - b)^2 + c$. For this reason any function that can be thrown into the form $a(x - b)^2 + c$ may be called a *parabolic function* of x.

Consider the functions $-2x^2 - 12x + 4$ and $+3x^2 - 12x - 3$. The first can be written $-2(x + 3)^2 + 22$ and the second $+3(x - 2)^2 - 15$. Both are, therefore, parabolic functions. The graph of the first is the "head up" parabola $y = -2x^2$ moved 3 units to the left and 22 units upwards; that of the second the "head down" parabola $y = +3x^2$ moved 2 units to the right and 15 units downwards. Moreover it is clear that this reshaping of the function could be carried out in the same way whatever numbers, positive or negative, replaced the -2, -12, and $+4$ of the former function. That is, a function of the form $px^2 + qx + r$ is always parabolic whatever the values of p, q and r.

We have seen that in a parabolic function a value of y can be found corresponding to any value of x that can be specified. But the converse statement does not hold. Thus the "head down" parabola

$$y = + \tfrac{1}{4}(x + 5)^2 + 7$$

of § 2 has no points lower than $y = +7$, and the "head up" parabola

$$y = - \tfrac{1}{4}(x - 10)^2 - 8$$

no points higher than $y = -8$. That is, the function $+\frac{1}{4}x^2 + \frac{5}{2}x + 13\frac{1}{4}$ has no values lower than $+7$ and the function $-\frac{1}{4}x^2 + 5x - 33$ none higher than -8. Similarly the parabola

$$y = - 2(x + 3)^2 + 22$$

has no points above $y = +22$, and the parabola

$$y = +3(x-2)^2 - 15$$

none below -15. Hence no value of x can give the function $-2x^2 - 12x + 4$ a higher value than $+22$, or the function $+3x^2 - 12x - 3$ a lower value than -15.

These results can easily be deduced algebraically. Consider the function $3x^2 - 12x - 3$ or $+3(x-2)^2 - 15$. Let x begin by being negative and very large, and let it approach zero. Then $x - 2$ will be negative all the time, but $(x-2)^2$ will, of course, be positive. Both statements remain true as x passes through zero and approaches $+2$. At this point $x - 2 = 0$ and $y = -15$. As x moves towards higher positive values $x - 2$ becomes positive and $(x-2)^2$ is, of course, also positive. Thus $(x-2)^2$ if it has a value at all is always positive. That is, it always makes the value of the function higher than -15. As x approaches and passes through $+2$, y descends to -15 and then ascends again. For this reason -15 is called the *turning value* of the function. In this case it is a *lower* turning value, because the function, after descending the number scale to -15, begins to ascend. Similar considerations show that, in the function

$$-2(x+3)^2 + 22$$

$(x+3)^2$ is always positive, being a square, and that $-2(x+3)^2$ is always negative. The value of the function is, therefore, always below $+22$ if $x + 3$ has any value at all. When $x = -3$ it has no value and the function has the *upper turning value* $+22$. It is obvious that a function whose graph is a "head up" parabola always has an upper turning value and one whose graph is "head down" always a lower turning value. The turning value is simply the distance the parabola has been moved up or down from what may conveniently be called the "standard position" in which its vertex coincides with the origin. Start, for example, with the head down parabola $y = +2{\cdot}3x^2$. Move it $3{\cdot}7$ to the right and $6{\cdot}5$ upwards and we have $y = +2{\cdot}3(x - 3{\cdot}7)^2 + 6{\cdot}5$. The function $2{\cdot}3(x - 3{\cdot}7)^2 + 6{\cdot}5$ has a lower turning value $+6{\cdot}5$ when $x = +3{\cdot}7$. Move the original parabola $4{\cdot}8$ to the left and $9{\cdot}3$ downwards and we have $y = +2{\cdot}3(x + 4{\cdot}8)^2 - 9{\cdot}3$. The function $2{\cdot}3(x + 4{\cdot}8)^2 - 9{\cdot}3$ has a lower turning value of $-9{\cdot}3$ when $x = -4{\cdot}8$.

[Ex. XLII, B, can now be worked.]

§ 4. *The Method of Differences.*—Given the graph of a function it is easy to find whether it is parabolic and if it is to determine its precise form. Draw any number of ordinates to the curve separated by any constant distance h. Measure their heights and set the results down in order in a column under the heading "y". Subtract the height of each ordinate from that of the next one to the right and set the differences down in another column. This column is called the "first differences of y" and is usually headed by the symbol $\Delta_1 y$. Next subtract each first difference from the one below it in the column and set these differences down in a third column. The third set of numbers is called the "second differences of y" and the column is headed $\Delta_2 y$. Now it is easy to show that if the curve is parabolic—(that is, if the function is of the form $px^2 + qx + r$)—the second differences should be constant to the degree of exactness which the means of measurement permit. Consider any three consecutive members of the series of equidistant ordinates and call the abscissa of the first x, so that the abscissæ of the other two will be respectively $x + h$ and $x + 2h$. Then the numbers in the three columns must be of the following form :—

y	$\Delta_1 y$	$\Delta_2 y$
$p^2x + qx + r$	$p(2xh + h^2) + qh$	$2ph^2$
$p(x + h)^2 + q(x + h) + r$	$p(2xh + 3h^2) + qh$	
$p(x + 2h)^2 + q(x + 2h) + r$		

But since the second difference, $2ph^2$, does not contain x it would obviously be the same whichever three consecutive ordinates had been taken. It follows that however many equidistant ordinates were measured the second differences all along the curve would have the constant value $2ph^2$. It is easy to see also that the second differences will not be constant *unless* the function is parabolic. For suppose it to be linear, that is of the form $px + q$. Then the height of the ordinate to the right of the one whose abscissa is called x is $p(x + h) + q$ and the first difference is ph. Thus, if the function is linear, the first differences are already independent of x and therefore will be constant. It follows that the second differences must be zero. These results are, of course, perfectly obvious from the consideration of a sloping straight line. On the other hand, if the function is of the third degree, that is, if it has the form

$$px^3 + qx^2 + rx + s,$$

it is evident that the first differences will all contain the term x^2. For example, the difference between the ordinates whose abscissæ are x and $x + h$ will begin with the difference between $p(x + h)^3$ and px^3. It follows from what was said above that we cannot now possibly reach constancy until we come to the *third* differences. A table, made on the same plan as the one above but involving four consecutive ordinates, will show that, as a matter of fact, the third differences are in this case constant and have the value $6ph^3$. By the use of Stifel's Table it is possible to show in exactly the same way that if the function is of the nth degree (where n is any given positive integer) the nth differences of the ordinates will be the first to be constant.

The method illustrated in this article is called *the method of differences.*

[Ex. XLII, C, can now be worked.]

C. Quadratic Equations (ch. XXIV., § 4; Ex. XLIII).

§ 1. *Graphic Determination of Values of x.*—The graph of a parabolic function shows that although to every value of x there corresponds only one value of y yet if a given value of y be specified there are two values of x which will yield it. In the graph these values are shown by drawing a horizontal line across the curve at the proper height or depth. For example, to find the values of x for which the function $\frac{1}{2}x^2 - 6x + 10$ has the value $+ 4{\cdot}5$ we must draw across the parabola[1] $y = \frac{1}{2}x^2 - 6x + 10$ the horizontal line $y = + 4{\cdot}5$. It will be found to cross the graph at the points where $x = + 1$ and $x = + 11$.

To find the values of x for which the value of the function becomes $- 6$ the line $y = - 6$ must be drawn cutting the graph where $x = + 4$ and $x = + 8$. As further examples note that $y = + 10$ is given by $x = 0$ and $x = + 12$; and that $y = + 16{\cdot}5$ is given by $x = - 1$ and $x = + 13$.

§ 2. *Solution by Calculation.*—Can these results be reached by calculation? To this question Ex. XLII, No. 17, suggests an answer. The values of x for which a parabolic function has zero value were there found by expressing the function as a product of two linear functions. In the present case. for example, we have :—

[1] Since $y = \frac{1}{2}x^2 - 6x + 10 = \frac{1}{2}(x - 6)^2 - 8$ this is the parabola $y = \frac{1}{2}x^2$ moved 6 units to the right and 8 downwards.

$$\begin{aligned} y &= \tfrac{1}{2}x^2 - 6x + 10 \\ &= \tfrac{1}{2}(x^2 - 12x + 20) \\ &= \tfrac{1}{2}\{(x - 6)^2 - 16\} \\ &= \tfrac{1}{2}(x - 2)\,(x - 10) \end{aligned}$$

It is now obvious that $y = 0$ both if $x = +2$ and if $x = +10$. This information can, then, always be obtained by calculation as well as by consulting a graph.

The same method is easily extended to the problems of § 1. For example, to find the values of x for which $y = +4{\cdot}5$ we need only lower the parabola through 4·5 units making the corresponding relation become

$$\begin{aligned} y &= \tfrac{1}{2}x^2 - 6x + 5{\cdot}5 \\ &= \tfrac{1}{2}\{(x - 6)^2 - 25\} \\ &= \tfrac{1}{2}(x - 1)\,(x - 11) \end{aligned}$$

The former line $y = +4{\cdot}5$ now coincides with the x-axis; and the factorized function shows that it cuts the parabola where $x = +1$ and $x = +11$. Since the movement of the curve was entirely vertical the line $y = +4{\cdot}5$ must have cut the original parabola in points having the same abscissæ. In other words the function $\frac{1}{2}x^2 - 6x + 10 = +4{\cdot}5$ when $x = +1$ and again when $x = +11$. Substitution (as well as inspection of the graph) shows this conclusion to be correct.

To find what values of x make $y = -6$ the parabola must be raised 6 units so as to bring the line $y = -6$ into coincidence with the x-axis. In the new position we have :—

$$\begin{aligned} y &= \tfrac{1}{2}x^2 - 6x + 16 \\ &= \tfrac{1}{2}\{(x - 6)^2 - 4\} \\ &= \tfrac{1}{2}(x - 4)\,(x - 8) \end{aligned}$$

In this case the x-axis cuts the curve where $x = +4$ and $x = +8$. These are, then, the values of x for which the original function has the value -6.

The other cases can be dealt with similarly. For $y = +10$ we must lower the parabola 10 units, for $y = +16{\cdot}5$, 16·5 units. We have, respectively :—

$$\begin{aligned} y &= \tfrac{1}{2}x^2 - 6x \\ &= \tfrac{1}{2}x(x - 12) \end{aligned} \qquad \text{and} \qquad \begin{aligned} y &= \tfrac{1}{2}x^2 - 6x - 6{\cdot}5 \\ &= \tfrac{1}{2}\{(x - 6)^2 - 49\} \\ &= \tfrac{1}{2}(x + 1)(x - 13) \end{aligned}$$

We conclude that, in the original function, $y = +10$ is given by $x = 0$ and by $x = +12$, and $y = 16{\cdot}5$ is given by $x = -1$ and by $x = +13$.

Next let us inquire for what values of x the function

$-\frac{1}{2}x^2 + 6x - 10$ has the value $+ 6$. Throwing the function into the form $-\frac{1}{2}(x - 6)^2 + 8$ we see that its graph is the former one inverted, for it is the parabola $y = -\frac{1}{2}x^2$ moved 6 units to the right and 8 units *upwards*. Two modes of solution are, therefore, possible. (*a*) The parabola may be lowered 6 units so that the line $y = + 6$ coincides with the x-axis and the function becomes :—

$$y = -\frac{1}{2}(x - 6)^2 + 2$$
$$= -\frac{1}{2}(x - 4)(x - 8)$$

which gives $y = 0$ for $x = + 4$ and $x = + 8$. (*b*) The parabola may be revolved about the x-axis. The corresponding function then becomes $y = \frac{1}{2}(x - 6)^2 - 8$ and the line $y = + 6$ becomes $y = - 6$. The problem is thus made identical with one solved above.

Lastly, if we ask (for instance) what values of x give $\frac{1}{2}x^2 - 6x + 10$ the value $- 8$ and $-\frac{1}{2}x^2 + 6x - 10$ the value $+ 8$, we see that the former graph must be raised and the latter lowered 8 units. In either case the vertex will now touch the x-axis at the point $x = + 6$. It seems, then, that there is one exception to the rule that a given value of a parabolic function always corresponds to two distinct values of x. But the algebraic solutions are :—

$$y = \tfrac{1}{2}x^2 - 6x + 18 \qquad y = -\tfrac{1}{2}x^2 + 6x - 18$$
$$= \tfrac{1}{2}(x - 6)^2 \qquad = -\tfrac{1}{2}(x - 6)^2$$
$$= \tfrac{1}{2}(x - 6)(x - 6) \qquad = -\tfrac{1}{2}(x - 6)(x - 6)$$

That is, the parabolic functions are factorizable, as before, into two linear functions, but these two are identical. It is best, therefore, to say that in such cases two identical values of x yield the given value of y.

§ 3. *Quadratic Equations.*—Consider the following batch of questions. For what values of x does :—

(i) the function $2x^2 - 12x + 9$ have the value $+ 23$;
(ii) the function $2x^2 - 12x - 5$ have the value $+ 9$;
(iii) the function $2x^2 - 12x - 17$ have the value $- 3$;
(iv) the function $-2x^2 + 12x + 27$ have the value $+ 13$;
(v) the function $-2x^2 + 12x - 3$ have the value $- 17$?

By § 2 the first three questions all reduce to the question : "Where does the parabola $y = 2x^2 - 12x - 14$ cross the x-axis?" and the last two to the question : "Where does the parabola $y = -2x^2 + 12x + 14$ cross the x-axis?" Moreover, since the latter parabola can be turned into the former by revolving it about the x-axis, the answer to the last two questions must

be the same as the answer to the first three. The answer to all five is, in fact, given by the argument:—

$$2x^2 - 12x - 14 = 0$$
$$x^2 - 6x - 7 = 0$$
$$(x - 7)(x + 1) = 0$$

Therefore either $x - 7 = 0$ or $x + 1 = 0$;
that is, either $x = +7$ or $x = -1$.

Such an argument may, then, be the means of solving any number of problems of the kind stated at the beginning of this article. For this reason it is important to become skilful in carrying out the process indicated. Any statement which, like the above, is of the form $px^2 + qx + r = 0$ is called a *quadratic equation* and the process just illustrated is called "solving the equation". The numbers which (like the -1 and $+7$ above) satisfy the equation are its *roots*. That is, the roots of a quadratic equation $px^2 + qx + r = 0$ are the values of x for which the parabolic function $px^2 + qx + r$ has zero value. To solve a quadratic equation we must express the parabolic function as a product of two linear functions. From these we form two simple equations. The roots of the two simple equations are the two roots of the quadratic equation. If the linear factors are identical the two roots are the same.

In the examples given above the parabolic function was generally factorized after being expressed as the difference between two squares. If the factors can be seen by inspection this step may be omitted. For instance, it is obvious that $x^2 - 7x + 12 = 0$ can be expressed as

$$(x - 3)(x - 4) = 0$$

and that the roots are $+3$ and $+4$. The longer method is, however, the best for difficult cases. It is also the only way to make sure whether the equation has roots or not. Take, for example, the equation $x^2 - 7x + 20 = 0$. Here we have

$$x^2 - 7x + 20 = (x - \tfrac{7}{2})^2 + 7\tfrac{3}{4}$$

but as the right-hand expression is the sum and not the difference of two squares the function cannot be factorized. The equation has, then, no roots. Consider, on the other hand, $x^2 - 7x + 9 = 0$. Here we have

$$x^2 - 7x + 9 = 0$$
$$(x - \tfrac{7}{2})^2 - \tfrac{13}{4} = 0$$

or $$\left(x - \frac{7}{2}\right)^2 - \left(\frac{\sqrt{13}}{2}\right)^2 = 0$$

whence $$\left(x - \frac{7 + \sqrt{13}}{2}\right)\left(x - \frac{7 - \sqrt{13}}{2}\right) = 0$$

Therefore either

$$x - \frac{7 + \sqrt{13}}{2} = 0 \text{ or } x - \frac{7 - \sqrt{13}}{2} = 0$$

That is, either

$$x = \tfrac{1}{2}(7 + \sqrt{13}) \text{ or } x = \tfrac{1}{2}(7 - \sqrt{13})$$

If these cases had not been dealt with in the way shown we could not have been certain that there are no roots in the first case and that there are roots, but complicated ones, in the second case.

[Exs. XLIII and XLIV may now be worked.]

D. Inverse Functions (ch. xxiv., § 6; Ex. XLV).

§ 1. *Changing the Subject of a Parabolic Formula.*—A ball is rolled up a smooth sloping board 75 inches long. Its velocity is at first 20 ins./sec. and falls off 4 ins./sec. every second. Find a formula for its distance from the top end of the board. The formula for the velocity is :—

$$v = 20 - 4t$$

therefore, as in Ex. XXX, Nos. 22 *et seq.*, we have, s being the distance from the starting point :—

$$\begin{aligned} s &= 20t - 4 \times \tfrac{1}{2}t^2 \\ &= 20t - 2t^2 \end{aligned}$$

If we put S for the distance from the farther end of the board we obtain :—

$$\begin{aligned} \mathrm{S} &= +75 - 20t + 2t^2 \\ &= +2t^2 - 20t + 75 \\ &= +2(t - 5)^2 + 25 \end{aligned}$$

The distance from the farther end of the board is, then, a parabolic function of the time and has a lower turning value of + 25. That is, the ball reaches a point 25 inches from the top of the board and then begins to descend again.

Now suppose we want to change the subject of this parabolic formula to t—that is, to obtain a formula giving the times at which the ball is at a given distance from the top of the board. (There will, in general, be two times for each

distance—one on the way up and one on the way down.) An easy way to proceed is as follows. Take the graph corresponding to the formula $S = + 2t^2$ and turn it into the position (fig. 75) in which t is measured along the vertical and S along the horizontal axis. It is now obvious that if we change the subject of the formula to t we shall obtain not one but two formulæ. The first is $t = + \sqrt{S/2}$ corresponding to the upper branch of the graph (OP, fig. 75); while the second is $t = - \sqrt{S/2}$ corresponding to the lower branch (OP'). But to make the new formulæ correspond to the original formula $S = 2(t - 5)^2 + 25$, i.e. $S - 25 = 2(t - 5)^2$, we must substitute $S - 25$ for S and $t - 5$ for t. In

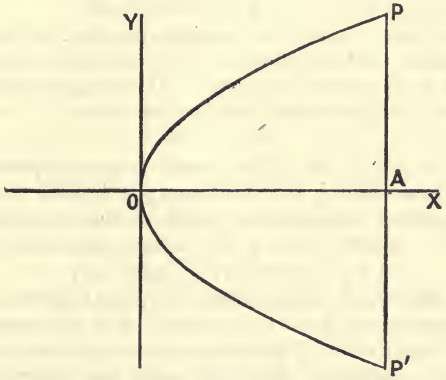


FIG. 75.

this way we arrive at the two formulæ :—

$$t - 5 = + \sqrt{(S - 25)/2}$$

or $$t = + 5 + \sqrt{(S - 25)/2}$$

and $$t = + 5 - \sqrt{(S - 25)/2}$$

The two formulæ may be summarized in the form :—

$$t = + 5 \pm \sqrt{(S - 25)/2}$$

The corresponding graph will be obtained by shifting fig. 75 to the right 25 units and 5 units upwards.

This combined formula shows once more that the ball cannot come within 25 inches of the top of the board. For if S were negative, or if, being positive, it were less than 25, the number under the radical sign would be negative. But

since a negative number has no square root there could be no corresponding value of t. That is, there is no time when the ball is within 25 inches of the top.

In general, if the original formula is :—

$$S = a\,(t - b)^2 + c$$

the formulæ with t as subject are :—

$$t = b \pm \sqrt{(S - c)/a}$$

In the above problem a is positive. To obtain a problem in which it is negative let us suppose that the distance of the ball is measured from a point 10 inches up the board from the bottom. Then we have :—

$$\begin{aligned} S &= s - 10 \\ &= -2t^2 + 20t - 10 \\ &= -2\,(t - 5)^2 + 40 \end{aligned}$$

This time there is an *upper* turning value of + 40—a result which agrees with the former lower turning value of + 25 measured to the top of the board. Beginning, as before, with the graph in the standard position, we have :—

$$S = -2t^2$$

and $t = \pm \sqrt{S/(-2)}$. This result is not impossible as it seems at first sight. It means simply that all the values of S are negative. Moving the graph so that it represents the relation $S = -2t^2 + 20t + 10$ we have the twin formulæ :—

$$t = +5 \pm \sqrt{(S - 40)/(-2)}$$

In order that the number under the radical sign may be positive the numerator must be negative. For this result S must either be negative or, if positive, numerically not greater than 40. These conditions obviously mean, once more, that S has an upper turning value of + 40.

In applying this method practically we proceed as follows :—

$$\begin{aligned} S &= a\,(t - b)^2 + c \\ (t - b)^2 &= (S - c)/a \\ t - b &= \pm \sqrt{(S - c)/a} \\ t &= b \pm \sqrt{(S - c)/a} \end{aligned}$$

If the original formula is given in the form

$$S = pt^2 + qt + r$$

we may, as an alternative method, write it

$$pt^2 + qt + (r - S) = 0$$

and by the formula of Ex. XLIV, No. 1, deduce at once that

$$t = \{-q \pm \sqrt{q^2 - 4p(r - S)}\}/2p$$

§ 2. *Inverse Functions.*—In the foregoing problem S was

a parabolic function of t. In other problems about moving bodies it might well be a linear function of the form $S = a + pt$ or a hyperbolic function of the form $S = k/(t + a) + b$. It is instructive to set side by side with these formulæ the corresponding formulæ in which t is the subject:—

$S = a + bt$	$t = (S - a)/b$
$S = k/(t + a) + b$	$t = k/(S - b) - a$
$S = a(t + b)^2 + c$	$t = -b \pm \sqrt{(S - c)/a}$

Each of the right-hand formulæ is said to be the *inverse* of the corresponding direct formula in the left-hand column. The relations between them are best brought out by replacing the concrete variables by the abstract variables x and y. In the direct formulæ t is the independent variable (to be represented by x); in the inverse formulæ the independent variable is S. We have therefore:—

Direct Functions.	*Inverse Functions.*
$y = a + bx$	$y = (x - a)/b$
$y = k/(x + a) + b$	$y = k/(x - b) - a$
$y = a(x + b)^2 + c$	$y = -b \pm \sqrt{(x - c)/a}$

The graphs of the inverse functions are most easily obtained by revolving the graph of the direct function through 180° about the x-axis and then turning it in its own plane, anticlockwise, through 90°. What were before values of x and y are thus converted into values of y and x. With the direct and inverse graphs before us the following points at once become clear: (*a*) Each of the direct functions is a *single-valued* function of x; that is, to every value of x there corresponds one and only one value of the function. (*b*) The inverse linear and the inverse hyperbolic functions are also single-valued, but the inverse parabolic function is *two-valued:* to each value of x there correspond *two* values of the function. In the direct and inverse linear functions the fields of x and y are both unlimited; that is, both variables are capable of assuming all values whatsoever. The same is true of the direct and inverse hyperbolic functions with the exception that when $x = -a$ in the former and $x = +b$ in the latter there is not really any corresponding value of y (A, § 1). But, in the direct parabolic function, while the field of x is unlimited that of y is unlimited one way only. If a is positive (that is, if the graph is "head down") y's value may be as high as you please but it can never be lower than c; if a is negative it may be as low as you please but

can never be higher than c. Finally, in the inverse parabolic function the opposite rule obtains. If a is positive $x - c$ must be positive so that x can never be lower than c; while if a is negative x can never be higher than c. On the other hand, the value of the radical may be (numerically) anything from zero upwards. Hence y may have any value as distant as we please above and below $-b$.

Consideration of the foregoing examples yields a simple rule for obtaining the inverse of a given function of x: Take the relation which expresses the original function; change its subject to x; in the result replace x by y and y by x. The relation so obtained is the expression of the inverse function.

[Exs. XLV and XLVI may now be worked.]

CHAPTER XXVII.

WALLIS'S LAW.

A. Area Functions (ch. XXIV., § 7 ; Ex. XLVII).

§ 1. *Area of the Parabola by Wallis's Method.*—The method of ch. XIX., B, can be used to calculate an area such as OPM in fig. 76, the curve being half of the parabola $y = kx^2$. Let $m + 1$ rectangles of equal breadth be set side by side, their areas being 0^2, 1^2, 2^2, . . . m^2. Behind them suppose $m + 1$ other rectangles (shown dotted in fig. 76), each equal to the largest of the former series, so that together they constitute a rectangle AM whose area is $(m + 1)m^2$. Then the increasing

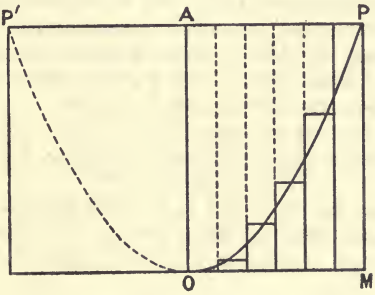


FIG. 76.

rectangles have a total area which is a certain fraction of AM, namely :—

$$\frac{0^2 + 1^2 + 2^2 + 3^2 + \ldots + m^2}{(m + 1)m^2}$$

The value of this fraction depends upon the value of m. If m assumes in succession the values 1, 2, 3, . . . we have :—

$$\frac{0+1}{2\times 1} = \frac{1}{2},\quad \frac{0+1+4}{3\times 4} = \frac{5}{12},\quad \frac{0+1+4+9}{4\times 9} = \frac{14}{36} = \frac{7}{18},\ \text{etc.},$$

the subsequent values being, in order,

$$\frac{3}{8'}\ \frac{11}{30'}\ \frac{13}{36'}\ \frac{5}{14'}\ \frac{17}{48'}\ \frac{19}{54'}\ \frac{7}{20'}\ \frac{23}{66'}\ \frac{25}{72'},\ \text{etc.}$$

A glance at these results suggests that they may be written :—

$$\frac{3}{6'}\ \frac{5}{12'}\ \frac{7}{18'}\ \frac{9}{24'}\ \frac{11}{30'}\ \frac{13}{36'}\ \frac{15}{42'}\ \frac{17}{48'}\ \frac{19}{54'}\ \frac{21}{60'}\ \frac{23}{66'}\ \frac{25}{72'},\ \text{etc.}$$

The law here followed is obviously given by the formula :—

$$\text{Fraction} = \frac{2m+1}{6m} = \frac{1}{3} + \frac{1}{6m}$$

It is difficult to suppose that this law, after holding good in the first twelve cases, should not continue to hold good. Assuming its truth for *any* number of rectangles we conclude that by making them numerous enough their combined area can be made to differ as little as we please from one-third of the area of AM. But as the rectangles are made thinner the area they cover will eventually become indistinguishable from the area under the curve OP. We conclude, therefore, that this area is—with an exactness more minute than can be measured by any number of decimal places—one-third of the rectangle AM. Since OM $= x$ and PM $= kx^2$ this conclusion may be expressed by the formula A $= \frac{1}{3}kx^3$.

It follows also that the area AOP $=$ AM $-$ OPM $= \frac{2}{3}$AM, and that the whole area of the parabola up to PP′ is

$$\tfrac{2}{3}\text{P}'\text{M} = \tfrac{2}{3}\text{OA} \times \text{PP}'.$$

This result can be expressed in another way by turning the original parabola into the form of fig. 75. The curve now corresponds to the relation $y = k'\sqrt{x}$ (where $k' = 1/\sqrt{k}$), and we have :—

$$\begin{aligned}\text{Area OPA} &= \tfrac{2}{3}\text{OA} \times \text{PA} \\ &= \tfrac{2}{3} \times x \times k'\sqrt{x} \\ &= \tfrac{2}{3}k'x\sqrt{x}\end{aligned}$$

§ 2. *Wallis's Law.*—The results of § 1 and of ch. XIX., B, can now be summarized as follows. Suppose an ordinate to start from the origin and to move to the right. If it has a constant height $y = k$ it will, in moving through a distance x, trace out an area A $= kx$. If its height is at first zero but increases in accordance with one of the laws $y = kx$, $y = kx^2$,

$y = k\sqrt{x}$, the area traced out will be given by the appropriate one of the corresponding laws $A = \frac{1}{2}kx^2$, $A = \frac{1}{3}kx^3$, $A = \frac{2}{3}kx\sqrt{x}$. Calling the function which gives the height of the ordinate the *ordinate-function*, and the function which gives the area traced out the *area-function*, the first three of these four results can be reduced to a very simple rule. For to the ordinate-functions :—

$$kx^0 \qquad kx^1 \qquad kx^2$$

correspond the area-functions :—

$$\tfrac{1}{1}kx^1 \qquad \tfrac{1}{2}kx^2 \qquad \tfrac{1}{3}kx^3$$

That is, if the ordinate-function is kx^{n-1} the area-function is $\frac{1}{n}kx^n$. Now John Wallis who discovered this rule (*Arithmetica Infinitorum*, 1655) was so struck with its simplicity that he sought to bring the fourth result under it also. His argument was practically as follows. The indices of the square roots of x^2, x^4, x^6, x^8, etc., are all formed by taking one half of the indices 2, 4, 6, 8, etc. It is natural, therefore, to inquire whether the square root of x^1 cannot, in accordance with the same rule, be expressed as $x^{\frac{1}{2}}$. Consideration shows that to do so would not contradict the law of indices. For, according to that law, if the symbolism $x^{\frac{1}{2}}$ can be permitted at all, it must have such a meaning that $x^{\frac{1}{2}} \times x^{\frac{1}{2}} = x^{\frac{1}{2}+\frac{1}{2}}$, i.e. $= x$. But this is precisely what it was intended to mean. We may, therefore, write $x^{\frac{1}{2}}$ for $\sqrt{x}$ and may, further, write

$$x\sqrt{x} = x \times x^{\frac{1}{2}} = x^{\frac{3}{2}}.$$

All four of our results now fall under the one rule, for when the ordinate-functions are :—

$$kx^0 \qquad kx^{\frac{1}{2}} \qquad kx^1 \qquad kx^2$$

the corresponding area-functions are :—

$$\tfrac{1}{1}kx^1 \qquad \tfrac{1}{\frac{3}{2}}kx^{\frac{3}{2}} \qquad \tfrac{1}{2}kx^2 \qquad \tfrac{1}{3}kx^3$$

The rule that to an ordinate-function kx^{n-1} there corresponds an area-function $\frac{1}{n}kx^n$ may be called Wallis's Law in memory of its discoverer. It has been shown to be true when n is 0, $\frac{1}{2}$, 1, 2. Later inquiries will determine whether it holds good also for other values of n.

§ 3. *Proof of the Theorem of* § 1.—In § 1 it was assumed that the fraction $(\Sigma m^2)/(m+1)m^2 = 1/3 + 1/6m$ in all cases because there seemed no reason why a law exemplified by twelve successive numerators and denominators should not

continue to hold good. This kind of reasoning is called *induction* and is obviously not completely satisfactory. A more satisfactory treatment would be one which *proves* that the law if true in one case will be true for all cases.

Let us suppose, then, that the rule holds when m has a certain value p—that is, when there are $p + 1$ rectangles of which the largest has an area p^2. Then we can show that it will hold good also when $m = p + 1$—that is, when there are $p + 2$ rectangles with $(p + 1)^2$ as the area of the largest.

By hypothesis, when there are $p + 1$ rectangles the fraction

$$\frac{0^2 + 1^2 + 2^2 + \ldots + p^2}{(p + 1)p^2} = \frac{2p + 1}{6p}$$

Hence $$0^2 + 1^2 + 2^2 + \ldots + p^2 = \frac{2p + 1}{6p} \times (p + 1)p^2$$

$$= \tfrac{1}{6}p(p + 1)(2p + 1)$$

Now let the number of rectangles be increased to $p + 1$. Then the numerator of the fraction will be

$$0^2 + 1^2 + 2^2 + \ldots + p^2 + (p+1)^2 = \tfrac{1}{6}p(p+1)(2p+1) + (p+1)^2$$

$$= \tfrac{1}{6}(p+1)(p+2)(2p+3)$$

But the denominator of the new fraction should now be $(p + 2)\,(p + 1)^2$. Dividing each side of the equality by this product we have

$$\frac{0^2 + 1^2 + 2^2 + \ldots + (p + 1)^2}{(p + 2)(p + 1)^2} = \frac{2p + 3}{6(p + 1)}$$

$$= \frac{1}{3} + \frac{1}{6(p + 1)}$$

We conclude that if the rule for the value of the fraction holds good when $m = p$ it holds good when $m = p + 1$. But the rule has been proved to hold good when m is any one of the numbers 1, 2, 3 . . . 12. Therefore it holds good universally.

This kind of argument is called a *proof by recurrence.*[1]

[Ex. XLVII may now be worked.]

B. Differential Formulæ (ch. XXIV., § 7; Exs. XLVIII, XLIX).

§ 1. *Differential Formulæ.*—In Ex. XLII, C, we saw how to determine the nature of a given graph by the "method

[1] Also *mathematical induction*—a bad term since it falsely suggests a similarity with the induction of § 1.

of differences". A series of equidistant ordinates is drawn, starting from any point on the graph. The heights of the ordinates are measured and a table is made of their first, second, third, . . . differences. We found that if the graph is a straight line the first differences will be constant, if a parabola, the second differences, while if it represents a function of x of the third degree the third differences are constant, and so on. We showed also that the converses of these statements are true, so that by examining the differences of the ordinates we can determine the "degree" of the function which corresponds to the graph.

The formulæ used in Ex. XLII, C, hold good if h, the distance between the ordinates, has *any* constant value. We are now to inquire what these formulæ become when h is taken so small that only those terms need be retained in which it appears in its lowest power.

It will be convenient to begin with the familiar case of the parabola $y = px^2 + qx + r$. With regard to the curves which correspond to formulæ of this form it was found (ch. XXVI., B, § 4) that

$$\text{the 1st diff. of } y = p(2xh + h^2) + qh = (2px + q)h + ph^2 \quad . \quad . \quad \text{(i)}$$

$$\text{the 2nd diff. of } y = 2ph^2 \quad . \quad . \quad . \quad . \quad \text{(ii)}$$

Now in (i) the second term has to the first the ratio

$$ph^2/(2px + q)h = ph/(2px + q) \quad . \quad . \quad \text{(iii)}$$

No matter, therefore, how large $|p|$ [1] or how small $|x|$ and $|q|$ may be, by taking h small enough this ratio may be made smaller than any given fraction. That is, the error produced by omitting the term ph^2 may be made as unimportant as we please. If we wish our formula to give results true within c per cent, we have only to reduce h below the value which makes the ratio in (iii) less than $c/100$. Provided that the value of h is less than this it is called "small," no matter what its absolute value may be. This rule is, in fact, to be regarded as the definition of the word "small" when it is applied to numbers in mathematics.

When the distance between the ordinates is meant to be "small" in this technical sense, it is convenient to denote the fact by replacing h by a special symbol. For this purpose the symbol δx is commonly employed. It must be carefully noted that the Greek letter delta is here the "grammalogue"

[1] The symbolism $|p|$ means "the numerical value of p".

not of the description of a number (as the x is) but of some such phrase as "a small increment in the value of". For brevity the whole symbolism δx may be read "the differential of x," the word "differential" being itself defined as meaning "a small increment in the value of". Similarly the difference between two ordinates whose distance apart is δx may be symbolized by δy—to be read "the differential of y". The second difference between the ordinates may conveniently be represented by $\delta^2 y$—to be read "the second differential of y," and so on. Finally we may note that although h^2 (when h is "small") should properly be written $(\delta x)^2$ it is usual to write it in the more compact form δx^2. No doubt the correct interpretation of δx^2 should be "the differential of the square of x". If it should prove necessary to symbolize this phrase the form $\delta(x^2)$ may be employed.

We are now in a position to express relations (i) and (ii) in the case when the distance between the ordinates is "small". The new formulæ will be

$$y = px^2 + qx + r \quad . \quad . \quad . \quad \text{A}$$
$$\delta y = (2px + q) \, . \, \delta x \quad . \quad . \quad . \quad \text{B}$$
$$\delta^2 y = 2p \, . \, \delta x^2 \quad . \quad . \quad . \quad . \quad \text{C}$$

Formulæ like B and C will be called *differential formulæ*. B (whose subject is the first difference of y) will be called a differential formula of the *first order*, C (whose subject is the second difference of y) one of the *second order*, and so on. It is important to remember that in these formulæ the symbol " = " asserts not absolute, but only approximate equality.[1] One of the advantages of using the special symbol δ in such a formula is that it may be taken as qualifying the meaning of the " = ," and so making it unnecessary to indicate in any other way that the statement *may* be true only to a certain degree of accuracy (Ex. IX, note to No. 8). Formula A, which contains no differentials and asserts exact equality, will be called (in relation to B and C) the *primitive* formula.

§ 2. *The Return from the Differential to the Primitive Formula.*—The investigation of § 1 illustrates the way in which differential formulæ may be derived from their primitive. The results may now be used to investigate the con-

[1] I.e. in some cases the equality may actually be exact but the formula only *guarantees* a certain percentage of accuracy.

verse problem—that of retracing the steps from a differential formula towards its primitive. For this purpose the three formulæ may more conveniently be set down in the forms

$$y = px^2 + qx + r \quad . \quad . \quad \text{A}'$$
$$\delta y/\delta x = 2px + q \quad . \quad . \quad . \quad \text{B}'$$
$$\delta^2 y/\delta x^2 = 2p \quad . \quad . \quad . \quad . \quad \text{C}'$$

Suppose, then, that we are given the relation C′ and are asked to find B′, how shall we proceed? The right-hand side of B′ is the sum of two terms, $2px$ and q. Of these the former is easily obtained, for it is merely the product of (the given) $2p$ by x. But whence are we to obtain the value of q? The answer is that from the information supplied (namely that $\delta^2 y/\delta x^2 = 2p$) we simply cannot tell what it is. All we can do is to set down the symbol q to indicate that the right-hand side of B′ may contain *some* added constant though we do not know *what* constant.

The source of this ambiguity may be made plain by numerical instances. Take the four following primitive formulæ:—

$$\left.\begin{array}{ll} y = 3x^2 & y = 3x^2 + 8 \\ y = 3x^2 + 5x + 1 & y = 3x^2 - 7x + 2 \end{array}\right\} \text{A}''$$
$$\left.\begin{array}{ll} \delta y/\delta x = 6x & \delta y/\delta x = 6x \\ \delta y/\delta x = 6x + 5 & \delta y/\delta x = 6x - 7 \end{array}\right\} \text{B}''$$
$$\left.\begin{array}{ll} \delta^2 y/\delta x^2 = 6 & \delta^2 y/\delta x^2 = 6 \\ \delta^2 y/\delta x^2 = 6 & \delta^2 y/\delta x^2 = 6 \end{array}\right\} \text{C}''$$

Now it is clear that, although the four parabolic functions in A″ agree only in having the same value for p (namely 3) and differ in respect of the values of q and r, these differences disappear by the time we reach C″. The formula $\delta^2 y/\delta x^2 = 6$ is determined simply and solely by the fact that $p = 3$, and, therefore, can give us information only about that constant. Any other parabolic function would yield the same differential formula of the second order if it also began with the term $3x^2$. In passing back from $\delta^2 y/\delta x^2 = 6$ to the formula for $\delta y/\delta x$, we can, therefore, only write $\delta y/\delta x = 6x + q$. This form implies that there *may* be a constant but that there is no evidence as to whether it is $+ 5$ or $- 7$ or some other of an endless range of possibilities.

Similarly if we are given B′ (that is, if p and q have assigned values) the passage back to A′ is only partly determinate. The term px^2 can be derived from $2px$ by multiplying by $x/2$, and the term qx from q by multiplying by x, but we have

no means of deciding the value of the constant r. We can only append the term $+r$ to the expression px^2+qx to indicate that there may be such a constant although we do not know what it is. Thus from the four differential formulæ B″ we can derive only the information

$$y=3x^2+r,\ y=3x^2+r,\ y=3x^2+5x+r,\ y=3x^2-7x+r.$$

Finally, if we are given $\delta^2y/\delta x^2=6$ and are asked to find the primitive, the introduction of an undeterminable constant will occur at each of the two stages in the backward process. From the given formula we derive in the first place the differential formula of the first order

$$\delta y/\delta x = 6x + q$$

and from this the primitive

$$y = 3x^2 + qx + r.$$

Of the numbers q and r we can only say that they may be zero or may have any values whatever.

§ 3. *Geometrical Meaning of the Constants.*—It is helpful to consider from the graphic point of view this appearance of an undeterminable constant at each step of the regress from a differential formula to its primitive. Take the primitive formula $y = 2x + 7$ which represents a certain straight line. Substituting $x + h$ for x and subtracting we have

the 1st difference of $y = 2h$

and, therefore,[1] $\delta y = 2 \,.\, \delta x$ or $\delta y/\delta x = 2$.

In this case the differential formula is determined by the coefficient of x in the primitive and by that alone. Any primitive of the form $y = 2x + q$ would yield the same differential formula $\delta y/\delta x = 2$. Conversely, from $\delta y/\delta x = 2$ we can deduce the primitive only in the ambiguous form $y = 2x + q$ where q may have any value including zero. Expressed graphically this means that from $\delta y/\delta x = 2$ we can deduce only that the primitive represents a straight line making with the x-axis the angle whose tangent is 2 but cannot deduce the position of the line. In other words, we can deduce that in the " standard position," in which it passes through the origin, the straight line would have $y = 2x$ as its formula, but we cannot deduce how far (if at all) the line has been raised or lowered. The differential formula may, therefore, be regarded

[1] Since there are no higher powers of h to neglect the equality is, on this occasion, absolute (see p. 284, footnote).

as a general description which includes in its reference $y = 2x$ and all possible lines parallel to it.

Similarly from the differential formula of the second order $\delta^2y/\delta x^2 = 6$ we can deduce that the primitive describes a parabola which in its standard position corresponds to $y = 3x^2$, but we cannot tell how much the parabola has been (i) raised or lowered, (ii) moved to the right or left. The differential formula is, therefore, a general description of all parabolas which can be derived from $y = 3x^2$ by these movements. On the other hand from the differential formula of the first order, $\delta y/\delta x = 6x + 5$, we can deduce not only that the parabola in its standard position would correspond to $y = 3x^2$, but also that it has been moved $\frac{5}{6}$ unit to the left. What remains uncertain is whether or how much it has been moved up or down. That is, $\delta y/\delta x = 6x + 5$ may be regarded as a general description of all parabolas which can be derived from $y = 3(x + \frac{5}{6})^2$ by vertical displacement.

§ 4. *Wallis's Law applies to Differential Formulæ.*—A differential formula of given order can always be derived from its primitive by the method of § 1. Nevertheless, it is profitable to inquire whether there are rules for writing down the results of the process without actually carrying it out. Such rules would be especially useful when we have to return from a differential formula towards the primitive, for the solution of this problem depends upon knowledge of the changes in the terms which accompany the transition from the primitive to the differential formula.

The results of § 1 suggest definite rules of this kind. In the first place it will be observed that qx in A′ becomes q in B′ and that $2px$ in B′ becomes in like manner $2p$ in C′. These results are similar because, although they appear at different stages of the transition from the primitive, they are produced by the same process—namely, that represented by the symbolism $k(x + h) - kx$. We have, therefore, the general rule (i) that in passing from a primitive to a differential formula of the first order, or from a differential formula of the first to one of the second order, and so on, any term kx becomes simply k. Conversely in taking any step backwards towards the primitive a constant k must be replaced by kx.

Again the term px^2 in A′ becomes $2px$ in B′ in virtue of the process $p(x + h)^2 - px^2$ followed by suppression of the term ph^2. It is clear that a term of the same form will undergo

the same transformation at any stage of the downward movement from the primitive formula. We have then the general rule (ii) that kx^2 becomes $2kx$ in a descending process, and that conversely kx becomes $\frac{1}{2}kx^2$ in an ascending process.

Lastly, the disappearance of the constant terms r in A′ and q in B′ in the descending process obviously exemplifies another general rule. The converse of this rule is the addition of an undetermined constant at each stage of the ascending process.

It will be noticed that rules (i) and (ii) are of the same form as Wallis's Law. That is, in the descending process a term of the form kx^n becomes nkx^{n-1}; in the ascending process kx^{n-1} becomes $\frac{1}{n}kx^n$. If this form is followed in other cases the rules for dealing with differential formulæ will be extremely simple. Let us then examine the case when $n = 3$. We have

$$\begin{aligned} y &= kx^3 \\ \text{1st difference of } y &= k(x+h)^3 - kx^3 \\ &= k(3x^2h + 3xh^2 + h^3) \\ \therefore\ \delta y &= 3kx^2 \,.\, \delta x \text{ or } \delta y/\delta x = 3kx^2. \end{aligned}$$

If we were given that $\delta y/\delta x = kx^3$ we should conclude by exactly the same argument that $\delta^2 y = 3kx^2 \,.\, \delta x^2$, or that $\delta^2 y/\delta x^2 = 3kx^2$, and similarly for any other step in the descending process. Thus this case is also covered by the form of Wallis's Law.

Suppose now that we are given a differential formula of the third order, of the form $\delta^3 y/\delta x^3 = p$; for example, $\delta^3 y/\delta x^3 = 5$. Then we have the following conclusions:—

$$\begin{aligned} \delta^3 y/\delta x^3 &= 5 \\ \delta^2 y/\delta x^2 &= 5x + p \\ \delta y/\delta x &= \tfrac{5}{2}x^2 + px + q \\ y &= \tfrac{5}{6}x^3 + \tfrac{1}{2}px^2 + qx + r \end{aligned}$$

p, q, and r being constants which may have any values, including zero.[1]

[1] The fact that three *independent* constants are introduced in the ascending process indicates that the primitive cannot in this case be simply a description of the curve $y = \frac{5}{6}x^3$ after a horizontal and a vertical movement. Ex. XLIV, Nos. 11-15, show why. By a suitable horizontal movement the graph of a cubic function of x which contains x^2 can always be turned into the graph of one which does not contain x^2. For example, $y = 2x^3 + 6x^2 + x - 2$ becomes $y = 2x^3 - 5x + 1$ when moved one unit to the right. The constant term

The question of the range of values of n over which Wallis's Law holds good must be left (as in ch. XXVII., A) for future discussion. There is, however, one more case of such importance in physical problems that it should be considered at once, namely, when $n = -1$. If the law is followed here kx^{-1} ought to become $-kx^{-2}$ in the descending process and kx^{-2} to become $-kx^{-1}$ in the ascending process. We have

$$\begin{aligned} y &= k/x \\ \text{1st diff. of } y &= k/(x+h) - k/x \\ &= -k \cdot \frac{h}{x(x+h)} \\ &= -k \cdot \frac{h}{x^2}\left(1 - \frac{h/x}{1+h/x}\right) \end{aligned}$$

Now, however small the numerical value of x, the value of h may be chosen so much smaller that the fraction h/x will become as small as we please. If x is positive $(h/x)/(1+h/x)$ is less than h/x and, therefore, by what has just been said, may also be made as small as we please. If x is negative $(h/x)(1+h/x)$ is greater numerically than h/x, but, again, by taking h small enough can be made as little so as we please. Thus in either case when h is small enough we may ignore the term $(h/x)/(1+h/x)$ and write

$$\delta y = -\frac{k}{x^2} \cdot \delta x \text{ or } \delta y/\delta x = -kx^{-2}.$$

Conversely in the ascending process a term kx^{-2} must be replaced by $-kx^{-1}$.

We have now shown, therefore, that Wallis's Law holds good in these descending and ascending transformations at least in the cases where $n = 1, 2, 3$ and -1.

§ 5. *The Meaning of the Inverse Process.*—We have postponed to the end what is in practice by far the most important of the questions we have to face: namely, the exact signi-

can next be removed by moving the curve (in its new position) vertically. Thus when $y = 2x^3 + 6x^2 + x - 2$ is moved one unit to the right and one unit downwards it becomes $y = 2x^3 - 5x$. But it is not possible in this way to remove the term $5x$ without reintroducing x^2 at the same time. That is to say, $y = 2x^3$ and $y = 2x^3 - 5x$ differ not only in position but also in shape. Thus from $\delta^3 y/\delta x^3 = p$ we cannot deduce even the shape of the curve. We can only deduce that it is one of the curves whose shapes are given by the formula $y = px^3/6 + qx$ where q may have any value.

ficance of the inverse process by which a primitive is deduced from a given differential formula. For example, supposing it to be given that $\delta y = 2x \,.\, \delta x$ what is the precise force of the conclusion that $y = x^2 + p$?

To answer this question we must remember that we are seeking what may be regarded as the unknown ordinate-function of a certain curve. The data suppose that equi-distant ordinates to this curve have been drawn and that the first differences between these ordinates are described, to a certain degree of accuracy, by the relation $\delta y = 2x \,.\, \delta x$. This degree of accuracy is fixed by the statement that even if the value of δy were really $2x \,.\, \delta x + a(\delta x)^2 + b(\delta x)^3 + \ldots$ (where a, b, etc., either are constant numbers or involve powers of x) the sum of the terms which follow the $2x \,.\, \delta x$ would be smaller than the smallest fraction of $2x \,.\, \delta x$ of which cognisance is to be taken. Let us suppose that H_o, the height of one of the ordinates, is known, and let its abscissa be x_o. Then if $2x \,.\, \delta x$ were the exact measure of the first difference the next ordinate, H_1, would be $H_o + 2x_o \,.\, \delta x$. Similarly H_2 would be $H_o + 2x \,.\, \delta x + 2x_1 \,.\, \delta x$; and, in general,

$$H_r = H_o + 2x_o \,.\, \delta x + 2x_1 \,.\, \delta x + \ldots + 2x_r \,.\, \delta x \quad . \quad \text{D}$$

δx being the constant distance between the ordinates and $x_1, x_2 \ldots$ being the abscissæ of H_1, H_2, etc. But the first differences are *not* exactly $2x_o \,.\, \delta x$, $2x_1 \,.\, \delta x$, $2x_2 \,.\, \delta x$, etc., but may depart from these numbers by any amount not greater than (say) c per cent. It follows, therefore, that the values of H_1, H_2, H_3, etc., calculated by formula D are not necessarily the exact heights of the ordinates but may differ from them by not more than c per cent. Since, however, by reducing δx, c may be made as small as we please, H_1, H_2, $H_3 \ldots$ can be calculated to any degree of accuracy.

There is to this procedure the obvious objection that if δx is small enough to secure a reasonable approximation to the value of H_r the number of additions must be unmanageably great. Consequently we are driven to inquire whether we cannot deduce from $\delta y = 2x \,.\, \delta x$ a formula which will enable us to calculate directly, if only approximately, any ordinate we please. At this point we recollect that if the formula of a curve were $y = x^2 + p$ the first difference would be given by the formula

$$\text{1st diff. of } y = 2xh + h^2$$
$$\text{or } \delta y = 2x \,.\, \delta x + (\delta x)^2 \text{ exactly.}$$

But by hypothesis δx is so small that it is indifferent whether

we equate δy to $2x \,.\, \delta x$ or to $2x \,.\, \delta x + (\delta x)^2$, for the added term is less than c per cent of the original $2x \,.\, \delta x$. Consequently it is legitimate to suppose that the first difference (which, by hypothesis, *may* be only approximately $2x \,.\, \delta x$) is actually $2x \,.\, \delta x + (\delta x)^2$. Upon this supposition it would follow at once that the ordinate-formula or primitive is $y = x^2 + p$.

To determine the amount of error to which this conclusion is liable we consider the most unfavourable case possible. By hypothesis, the true value of the difference *may* be c per cent less than $2x \,.\, \delta x$ and the value of $2x \,.\, \delta x + (\delta x)^2 c$ per cent more than $2x \,.\, \delta x$. Thus in assuming the first difference to be $2x \,.\, \delta x + (\delta x)^2$ we may be over-estimating it to the extent of about $2c$ per cent of its true value. Consequently all the ordinates calculated by the formula $y = x^2 + p$ *may* be about $2c$ per cent too high. But since the value of c depends merely upon the size of δx we can make it as small as we please by supposing δx to be small enough. That is to say, the formula $y = x^2 + p$ gives, if not an absolutely exact measure of the heights of the ordinates, at least as close an approximation as anyone can require.

As a second example suppose it to be given that $\delta y/\delta x = -1/x^2$ or $\delta y = -\delta x/x^2$. Here we must remember that when $y = 1/x$ the first difference of y is given by the formula

$$\text{1st diff. of } y = -\frac{h}{x^2}\left(1 - \frac{h/x}{1 + h/x}\right)$$

$$\text{or } \delta y = -\frac{\delta x}{x^2} + \frac{(\delta x)^2}{x^3(1 + \delta x/x)} \text{ exactly .} \quad \text{E}$$

As before, the second term is, by hypothesis, less than the smallest number of which cognisance is to be taken. To assume, therefore, that δy is given by the relation E does not contradict the datum that it is approximately $-\delta x/x^2$. On the other hand, that assumption enables us to conclude that the ordinates are given to any degree of accuracy anyone chooses to name by the formula $y = 1/x$.

It will be seen that similar reasoning holds good of the steps from the second differential formula to the first, etc. The conclusion of the whole argument is, therefore, that when a differential formula of any order holds good in the sense described in § 1 the primitive deduced from it by such rules as those of §§ 3, 4 also holds good to an unlimited degree of accuracy.

[Exs. XLVIII, XLIX, may now be worked.]

CHAPTER XXVIII.

THE CALCULATION OF π AND THE SINE TABLE.

A. Calculation of π (ch. XXIV., § 9 ; Ex. L).

§ 1. *The Principle to be used.*—Hitherto nothing has been said about the mode of calculating π. Its value can, of course, be determined only roughly by the measurement of actual circles. As early as 250 B.C. the great Archimedes reached by calculation the number $3\frac{1}{7}$ which we so frequently use, and his method was essentially the one we shall follow, although we will study it in a form due to Ludolph van Ceulen (1586).

Consider two regular polygons with the same number of sides respectively inscribed within and circumscribed without any circle. It is assumed as obvious (i) that the ratio of the perimeter of the inscribed polygon to the diameter of the circle is less, and that of the perimeter of the circumscribed polygon greater, than the ratio of the circumference of the circle to the diameter; (ii) that the difference between the first two of these ratios can always be made smaller by increasing the number of sides, but that neither of them can ever become exactly equal to the third.[1] It follows that by taking an inscribed and a circumscribed polygon of a sufficiently large number of sides and calculating the ratios of their perimeters to the diameter we can find a pair of numbers, as close to one another as we please, between which π must lie (cf. p. 94).

Taking polygons containing 96 sides Archimedes showed in this way that π lies between $\frac{22}{7}$ and $\frac{223}{71}$.

[1] No elementary student is likely to question these assumptions. Strictly speaking, however, they are really a definition of what we mean by "the length of circumference of the circle". See Young, *Fundamental Concepts of Algebra and Geometry*, p. 205.

§ 2. *Ludolph van Ceulen's Method.*[1]—Van Ceulen had two advantages over Archimedes: a system of symbols by which he could easily formulate the ratio corresponding to *any* number of sides, and the decimal fractional notation[2] in which the numerical value of a ratio could be expressed to any required degree of exactness. Armed with these instruments he sought in the following way to obtain a double series of numbers, like those of p. 94, between the successive pairs of which the value of π may be imprisoned within limits as narrow as the calculator pleases.

Let AB (*Exercises*, fig. 47) be any chord in a circle of unit radius, and let AB′ be a chord bisecting the arc AB. Let C = the length of the chord AB, and C_1 = the length of the chord AB′; also let OH = b, B′H = a, H being the foot of the perpendicular from the centre upon AB. Then since B′AC is a right angle, we have

$$\begin{aligned} C_1^2 &= a \times B'C \\ &= 2a \\ &= 2(1 - b) \\ &= 2\{1 - \sqrt{1 - (\tfrac{1}{2}C)^2}\} \\ &= 2 - \sqrt{(4 - C^2)} \end{aligned}$$

Hence $$C_1 = \sqrt{\{2 - \sqrt{(4 - C^2)}\}}$$

Next, bisect the arc AB′ at B″. Then putting C_2 for the length of the chord AB″, we have by the foregoing formula,

$$\begin{aligned} C_2 &= \sqrt{\{2 - \sqrt{(4 - C_1^2)}\}} \\ &= \sqrt{:} 2 - \sqrt{:} 2 + \sqrt{(4 - C^2)} \end{aligned}$$

the notation "$\sqrt{:}$" being used[3] to signify that the radical sign covers everything that follows it.

It is now possible to formulate the length of the chord obtained after any number of bisections of AB. It will be, in fact,

$$C_p = \sqrt{:} 2 - \sqrt{:} 2 + \ldots + \sqrt{:} 2 + \sqrt{(4 - C^2)}$$

the symbolism "$\sqrt{:}2$" being repeated p times.

If C is the side of an inscribed regular polygon of n sides then C_p is the side of one containing $N = n \times 2^p$ sides. The

[1] *De circulo et adscriptis liber*, 1619 (Latin translation by Snell of the Dutch original of 1596).

[2] He actually expresses his results as vulgar fractions but the denominators are always powers of 10. It should, however, be remembered that the Greeks used a system of "sexagesimals" not altogether unlike decimals (see *Exercises*, Pt. II, p. 97).

[3] Van Ceulen actually used "$\sqrt{.}$" in this way.

perimeter will be NC_p and the ratio of this number to the diameter, 2, that is $\frac{1}{2}NC_p$, will be a number which constantly approaches π as p is increased. The formula for the ratio to the diameter of the perimeter of the corresponding circumscribed polygon can be deduced from *Exercises*, fig. 48. Let PQ and AB be sides of the corresponding circumscribed and inscribed polygons. Put $AB = C_p$ and $PQ = E_p$. Then, by similar triangles,

$$\begin{aligned} E_p/C_p &= OT/OH \\ &= 1/b \end{aligned}$$

whence

$$\begin{aligned} E_p &= C_p/b \\ &= C_p/\sqrt{(1 - \tfrac{1}{4}C_p^2)} \\ &= 2C_p/\sqrt{(4 - C_p^2)} \end{aligned}$$

Whatever, then, be the length of the original chord C and whatever the value of p, it is certain that π lies between

$$\tfrac{1}{2}NC_p \text{ and } \tfrac{1}{2}NE_p = NC_p/\sqrt{(4 - C_p^2)}.$$

§ 3. *Actual Calculation of* π.—For the actual calculation of π we may start with any convenient chord AB. For example, if AB is the side of a square, then $C = \sqrt{2}$ and we have

$$\begin{aligned} C_p &= \sqrt{:}\,2 - \sqrt{:}\,2 + \sqrt{:}\,2 + \ldots + \sqrt{(4 - 2)} \\ &= \sqrt{:}\,2 - \sqrt{:}\,2 + \sqrt{:}\,2 + \ldots + \sqrt{2} \end{aligned}$$

If AB is the side of a hexagon $C = 1$ and the formula becomes

$$C_p = \sqrt{:}\,2 - \sqrt{:}\,2 + \sqrt{:}\,2 + \ldots + \sqrt{3}$$

With inexhaustible patience van Ceulen calculated the value of π by both methods—and by still others—to an enormous degree of minuteness. Thus starting with the inscribed square he calculated the perimeters of inscribed polygons of 8, 16, 32, . . . up to 1,073,741,824 sides. In the case of the last polygon he calculated also the perimeter of the corresponding circumscribed figure. In this way he found that the value of π lies between the numbers

3·1415926535897959

and 3·1415926535897932

Starting now with the hexagon, he calculated the perimeters of the inscribed and circumscribed polygons of 6,442,450,944 sides and hence determined that π lies between the numbers

3·141592653589793238

and 3·141592653589793239

It would be useless to repeat the details of these prodigious computations, but it is worth while to draw up a briefer table giving the perimeters of the inscribed and circumscribed polygons of 3, 6, 12, . . . sides until they are identical as far as the (nearest) fourth place of decimals. We may call it " Van Ceulen's Table ":—

N	$\frac{1}{2}$NC	$\frac{1}{2}$NE
3	2·5981	5·1962
6	3·0000	3·4641
12	3·1058	3·2154
24	3·1326	3·1597
48	3·1293	3·1461
96	3·1410	3·1427
192	3·1415	3·1419
384	3·1416	3·1416

We conclude that the value of π to four decimal places is 3·1416.

[Ex. L, A, can now be worked.]

B. The Calculation of the Sine-Table (ch. XXIV., § 9; Ex. L).

§ 1. *Preliminary.*—The seaman and the surveyor, to say nothing of the astronomer, require tables of sines, cosines, etc., much more exact than those obtainable by measuring triangles as in ch. XIII. We are to inquire, therefore, how a table of these ratios, carried to any given number of decimal places, can be computed. Note, first, that $\cos\alpha$ and $\tan\alpha$, and also $\cot\alpha$ (that is, $1/\tan\alpha$), $\sec\alpha$ (that is, $1/\cos\alpha$) and $\operatorname{cosec}\alpha$ (that is, $1/\sin\alpha$) can all be deduced from $\sin\alpha$. Thus the calculation of sines only will suffice. Moreover since $\sin(90^\circ - \alpha) = \cos\alpha$ and $\tan(90^\circ - \alpha) = \cot\alpha$ we need actually calculate only the sines from 0° up to 45° in order to have data for computing the whole of the tables.

§ 2. *The Calculation of Sines.*—Ch. XIV., B, showed how the sines could be calculated for angles of 60° and 45°. The method of Section A of this chapter can be used to deduce from these the sines of angles which are $\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{8}$, etc. of 60° or 45°. For if, in *Exercises*, fig. 47, the angle $AOH = \alpha$ it is evident that $AB = 2AH = 2\sin\alpha$. In fact, the earlier astronomers, such as Hipparchus (about 130 B.C.) and Ptolemy (about 120 A.D.) used chords where we use sines. Sines,

cosines, and tangents appear only at the end of the Middle Ages.

The values of C_p in van Ceulen's table (p. 295) are the chords subtended by angles of 60°, 30°, 15°, etc., in a circle of unit radius. The halves of these numbers are, therefore, the sines of 30°, 15°, 7·5°, 3·75°, 1·875°, 0·9375°, 0·46875°. Note that to four decimal places

$$\sin 0{\cdot}9375^\circ = 0{\cdot}0164 = 2 \times 0{\cdot}0082 = 2 \sin 0{\cdot}46875^\circ.$$

That is, within the first degree the sine is, to four decimal places, proportional to the angle. The value of sin 1° can, therefore, be calculated by proportion :—

$$\sin 1^\circ = \sin 0{\cdot}9375^\circ \times \frac{1{\cdot}0000}{0{\cdot}9375}$$
$$= 0{\cdot}0174$$

Now we know sin 15°, sin 30°, sin 45°. If, then, we had a formula for calculating sin $(\alpha + 1^\circ)$ when sin α and sin 1° are known we could, starting from either of these sines, and working backwards and forwards, calculate all sines up to 45°. Using sin $30' = \frac{1}{2}$ sin $1^\circ = 0{\cdot}008775$ we could similarly calculate the sines of the half-degrees, etc., etc. It is easy to find a formula which, given the sines and cosines of two angles, α and β, will enable us to calculate sin $(\alpha + \beta)$.

Let the circle in *Exercises*, fig. 49, have unit radius. Let AOB $= 2\alpha$ and BOC $= 2\beta$. Then our problem may be expressed thus: Knowing the lengths of AB and BC how can we calculate the length of AC? We have

AB $= 2 \sin \alpha$, BC $= 2 \sin \beta$, and AC $= 2 \sin (\alpha + \beta)$.

Also, since ACB at the circumference and AOB at the centre are on the same arc AB, ACB $= \frac{1}{2}$ AOB $= \alpha$. Similarly BAC $= \beta$. Thus

$$\text{AC} = \text{AB} \cos \beta + \text{BC} \cos \alpha$$

That is $$2 \sin (\alpha + \beta) = 2 \sin \alpha \cos \beta + 2 \sin \beta \cos \alpha$$

or $$\sin (\alpha + \beta) = \sin \alpha \cos \beta + \cos \alpha \sin \beta.$$

As an example take the calculation of sin 16° to four places. We have

$$\sin 15^\circ = 0{\cdot}2588 \qquad \sin 1^\circ = 0{\cdot}0174$$
$$\cos 15^\circ = \sqrt{(1 - \sin^2 15^\circ)} \qquad \cos 1^\circ = \sqrt{(1 - \sin^2 1^\circ)}$$
$$= 0{\cdot}9659 \qquad = 0{\cdot}9998$$
$$\sin 15^\circ \cos 1^\circ = 0{\cdot}2587$$
$$\cos 15^\circ \sin 1^\circ = 0{\cdot}0168$$
$$\therefore \quad \sin 16^\circ = 0{\cdot}2755$$

[Ex. L, B, can now be worked.]

SECTION III.

LOGARITHMS.

THE EXERCISES OF SECTION III.

*** The numbers in ordinary type refer to the pages of *Exercises in Algebra*, Part I; the numbers in heavy type to the pages of this book.

EXERCISE		PAGES
LII.	Growth Factors	269, 302
LII.	Growth Problems	272, 313
LIII.	The Gunter Scale	277, 319
LIV.	Logarithms and Antilogarithms	281, 325
LV.	The Base of Logarithms	283, 333
LVI.	Common Logarithms	286, 335
LVII.	The Use of Tables of Logarithms	289, 337
LVIII.	The Logarithmic and Antilogarithmic Functions	292, 341
LIX.	Nominal and Effective Growth Factors	297, 346

CHAPTER XXIX.

THE PROGRAMME OF SECTION III.

§ 1. *The Nature of Logarithms.*—Logarithms can be looked at from two distinct points of view. We may regard them (as their inventor did) as "artificial numbers" intended to facilitate computation, or we may regard them as the natural mode of expression of an important kind of connexion between variables. The usual method of teaching logarithms emphasizes the first point of view to the practical exclusion of the second; the method outlined in Section III attempts to combine the two in a treatment consonant with the principles followed in developing the previous sections. It begins with the study of certain concrete problems—here called "growth problems"—in which a particular kind of connexion between the variables is exemplified. So long as we consider only integral values of the independent variable (Ex. LI) these can be solved by ordinary arithmetic, but other cases are found to require a different method of solution. At first (Ex. LII) we are contented to deal with these refractory cases by means of graphs—here called "growth-curves"—but we soon pass to the consideration of more economical and accurate methods of procedure. After a transitional period (Ex. LIII) in which calculating devices (the "Gunter scale" and the "slide-rule") derived from the "growth-curves" are considered, we finally reach the notion of logarithms (Exs. LIV-LVI) and the use of tables (Ex. LVII). Then comes the usual second stage of the inquiry (Ex. LVIII) in which we turn our attention from the concrete problems and the methods needed to solve them, and direct it to the properties of the function which the behaviour of the concrete variables exemplifies. In connexion with this part of our investigation we find occasion for the first time to introduce the concept of an index as a number which may have any value, positive or negative,

integral or fractional. The section ends with a further study of practical problems from which there emerge the important notions connected by mathematicians with the algebraic symbol "e".

We have already advanced (ch. v., § 4) some reasons for ignoring both the historical accident that logarithms were invented in order to lighten the labour of computing products and quotients, and also the custom of deriving their theory from the theory of fractional indices. In brief, the plan here adopted of teaching logarithms as an alternative to the use of growth-curves in the solution of certain problems presents two advantages: in the first place it makes their practical value much more evident, and in the second place it prepares the way more effectively for the study of an extremely important type of function. In connexion with these arguments it is interesting to note that the calculation of logarithms was conceived by Napier himself as a kind of "growth-prob-

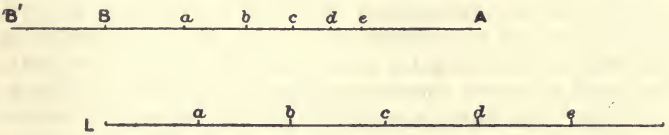


FIG. 77.

lem". This will be evident from the following summary of his explanation of the principle underlying his method.[1] In studying it the reader must remember that Napier was chiefly concerned to facilitate evaluations of formulæ in spherical trigonometry. For this reason his logarithms were the logarithms of sines. Moreover, the logarithm of the largest sine (= 1) was taken as zero, and the logarithms increased as the sine of the angle diminished. This device was adopted in order that in computing spherical triangles sin 90° (which occurs very often) might be ignored.

Let the line AB (fig. 77) be graduated uniformly from 0 to 1. Distances measured along this line from A may be taken as representing all the possible values assumed by the sine of an angle as the latter increases from 0° to 90°. Next sup-

[1] A translation (by W. R. Macdonald) of Napier's *Mirifici Logarithmorum Canonis Constructio* was published in 1889 by Blackwoods but appears now to be out of print.

pose a point to start from B and to move along the line towards the right with a speed always proportional to its distance from A. Let *a*, *b*, *c*, *d*, etc., mark points which it reaches after 1, 2, 3, 4, etc., equal intervals of time. Then it follows from the law of diminishing speed of the moving point that both the distances B*a*, *ab*, *bc*, *cd*, etc., and the distances BA, *a*A, *b*A, *c*A, etc., diminish in a constant ratio. It is obvious that the moving point will never quite reach A. While the first point is moving in the manner described along the line BA, let another point, starting with the same speed, move *uniformly* along the line L. Let *a*, *b*, *c*, *d*, on this line represent the points reached by the second moving point at the ends of the equal intervals of time which bring the first moving point to *a*, *b*, *c*, *d*, etc., on BA. Since these points will be equidistant the line which begins at L must be supposed of indefinite length towards the right. Consider any one of the points indicated on L—for example the fourth point, *d*. Then the equal segments L*a*, *ab*, *bc*, *cd*, correspond in succession to segments B*a*, *ab*, *bc*, *cd* on AB whose lengths have equal ratios. If we take the lengths of the equal segments of L as unity, we can say that the length of the line L*d* gives us the number of the distances in equal ratio necessary to bring the moving point from B to the corresponding point *d* on the line AB. The same thing is true of any of the points marked on L. For this reason Napier called the lengths 0, L*a*, L*b*, L*c*, etc., on the line L the *logarithms* (λογῶν ἀριθμος = number of the ratios) of the lengths AB, A*a*, A*b*, A*c*, A*d*, etc., on the line AB. But the points *a*, *b*, *c*, *d*, etc., on AB mark the position of certain numbers on the scale of sines. The corresponding lengths on the line L may then be called the logarithms of the sines.

Now set the line BA at right angles to L so that B coincides with L (fig. 78). Further, place the lengths A*a*, A*b*, A*c*, etc., at right angles to L at the points *a*, *b*, *c*, etc., in order. Then we have a series of ordinates whose lengths fall off in constant ratio as they recede uniformly from the origin. It follows that if we draw a smooth curve through the upper ends of the ordinates we obtain a "growth curve" of the kind studied in Ex. LII.

Napier did not himself pursue his argument to the consequence pointed out in the last paragraph. This step is said to have been first taken by Edmund Gunter (1581-1626).

At any rate it was Gunter who first conceived (1623) the idea of laying off the length of the ordinates of an exponential curve along its line of abscissæ and so produced a mechanical instrument which became (through Wingate, 1627, Oughtred, 1627, Milborne, 1650, and Partridge, 1657) the ancestor of the slide rule used so much by engineers to-day.

§ 2. *Growth Problems* (Exs. LI, LII; ch. xxx.).—The word "growth" as used in Section III scarcely needs formal definition. It refers to the familiar fact that measurable things of many kinds, when observed from time to time, are found to increase or decrease in magnitude. In some cases

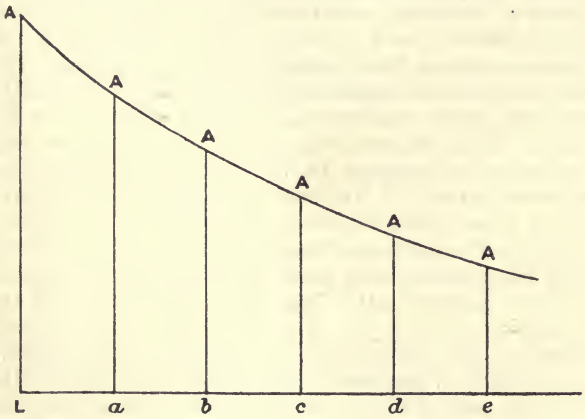


FIG. 78.

this growth is regular—that is, the magnitudes at different times succeed one another in accordance with a single mathematical law; in other cases it is irregular—that is, the changes during different periods follow different mathematical laws. "Regular" growth may conceivably take place in an endless number of different ways, but two types occur so frequently in the phenomena of nature and society that they deserve special consideration and study. The distinctive mark of the first type is that in equal times, however small, the magnitude of the growing thing always shows equal increments or equal decrements; that of the second is that the magnitudes at the beginning and the end of equal intervals of time are always

in the same ratio. The former law is characteristic of some of the simplest and most widespread phenomena of nature —for example, it is exhibited by the increasing velocity of a freely falling body or the decreasing velocity of a body moving against a constant resistance. It is also the law of "simple interest" and in this capacity plays an important part in economic affairs. The latter law is also exemplified in physical phenomena—for example, in the way in which the temperature of a hot body falls when cooling under constant conditions. But, just as physical phenomena give the most impressive examples of the operation of the law of constant increment or decrement, so biological and economic phenomena exemplify most strikingly the law of constant ratio. Not only is it the law of "compound interest" and in this capacity the final arbiter of the world of finance. In addition it is the law which, in its direct or derived forms, rules most of the phenomena of vital growth and decay and so is at once the law of life and of death. A method of dealing with the theory of logarithms which is based upon a careful study of this law seems sufficiently justified by that circumstance alone.

Ex. LI introduces the subject by examples in which growth is seen to be measured much more appropriately by a "growth-factor" than by the simple increment or decrement. Thus it is found in **No. 5** that the growth-factor of the height of the average American girl[1] is almost constant for several years, although the actual increment of her stature is never the same for two years in succession. It is obvious that a statement about the value of this constant factor gives much more concise and illuminating information than a statement about the values of the annual increments.

The examples of division A give practice in the application of the idea of a constant growth-factor. Only integral time-intervals are considered, and the investigation is summed up in **No. 11** in the formula $Q = Q_0 r^n$ where r is the growth-factor and n may have any integral value, positive or negative. In these examples the growth-factor is always given; in those of division B the problem is to calculate the growth-factor, given the magnitude of the growing thing at two moments separated by a specified integral number of time

[1] The table (*Exercises*, p. 269) is taken from Stanley Hall's *Adolescence*.

intervals. These examples should play an important part in preparing the student's mind for the arguments of the later exercises.

The tables asked for in **Nos. 13, 14** will be required in the next exercise. They should, therefore, be computed carefully and preserved.

Ch. xxx. begins with a review of the subject of regular growth in which some of the ideas of the present article are taught by simple examples. It goes on to develop a graphical method of solving "growth-problems" when the time is not an integral number of unit intervals. The argument is given fully in ch. xxx., but the teacher may be reminded in connexion with § 4 of the discussion in ch. iv., § 6.

The examples of Ex. LII are so framed that they can all be solved by "growth-curves" in which the factor is either 1·1, 1·25, or 1·3. These three curves are represented in *Exercises*, fig. 50, p. 273. The teacher may, of course, prefer to make his pupils use their own graphs.

§ 3. *The Gunter Scale; the Slide Rule* (Exs. LIII, LIV, A; chs. xxxi., xxxii., A). As was explained in § 1 the arguments of these exercises are intended to be a bridge between the solution of growth-problems by a graph and the use and theory of logarithms. The student who has worked a number of the examples contained in them will have already acquired the essential ideas underlying the use of logarithms. Moreover, these ideas should have for him the vividness and exactness which we have seen to be the peculiar gift of an appropriate graphical method of approach to abstract notions (ch. iv., § 1).

The special contribution of ch. xxxi. and Ex. LIII is the idea that *any* growth-curve, if prolonged far enough, supplies the means of solving any growth-problem, no matter what special factor may be involved in it. It shows, further, that the only thing actually needed is a record of the heights of the ordinates of such a curve distributed at proper intervals along a strip of paper. Such a graduated strip is here called a Gunter scale (see p. 302). The argument (ch. xxxi., § 3)—which shows that a curve need be used to graduate the Gunter scale only from 1 to 10, and that the graduations can then be extended indefinitely far to the right or left—is of great importance. The teacher will see that it prepares the student to recognize and understand the special value of logarithms

to base 10. Similarly the method of § 4 is not only interesting in itself but is also useful as a preparatory illumination of the idea that logarithms may be calculated to any base. This method may with advantage be demonstrated upon the squared blackboard with the aid of a long Gunter scale. A suitable scale is constructed by cutting a strip of "semi-logarithm paper" (supplied by instrument makers at threepence or fourpence a sheet) and sticking it parallel and near to an edge of a large sheet of ordinary drawing paper. A line of suitable length should be drawn upon the sheet at right angles to the strip, meeting it at the end where the graduations begin. Another straight line should be drawn to join the further ends of the line and the strip. The line may now be graduated similarly to the strip by the familiar method which consists in drawing lines parallel to the hypothenuse of the triangle from the graduations of the strip to meet the line at right angles to it. In this way a Gunter strip of any required length may be rapidly obtained.

Up to this point the properties of the growth-curve and the Gunter scale have been used only for finding powers and roots. In ch. XXXII., A, and Ex. LIV, A, the student learns that they can also be used for abbreviating the calculation of products and quotients. Once more an instrumental device is introduced as the best means of ensuring a thorough understanding of the uses of the table of logarithms. It is shown that multiplication and division may be performed with great ease by employing a second Gunter scale. In this way the student makes acquaintance with the slide rule. It will be observed that the examples are intended not to give practical skill in the use of the rule but to make the study of the rule illuminate the properties of logarithms. Apart from this pedagogical consideration there is no doubt some convenience in an early introduction of the slide rule. When its principles have been mastered in Ex. LIV, practical skill in using it may, at the option of the teacher of mathematics or his scientific colleague, be given by special exercises in computation. It must, however, be repeated that the rule, like the Gunter scale, is introduced here as an illustrative device intended to secure a really clear appreciation of the nature and properties of logarithms. It may in this capacity have full value for a student who never acquires facility in using it for computative work. If it is used merely as a teaching

aid there is no need to supplement the simple arrangements described in ch. XXXII., A. If it is to be used for actual computation it will be necessary to invoke the aid of the instrument maker and the guidance of a special text-book such as the excellent "Slide-Rule Notes" published in this series by Colonel Dunlop and Mr. C. S. Jackson.

§ 4. *Logarithms and Antilogarithms* (Exs. LIV, B, LV; chs. XXXII., B, XXXIII.).—The step from the Gunter scale to the table of logarithms is very short. Both may be regarded as derived from the growth-curve and as merely a convenient means of applying its properties. In the scale the graduations record the heights of the ordinates of the curve and the abscissæ are represented graphically by the distances of the graduations from the beginning of the scale. In the table the abscissæ are themselves recorded numerically and constitute the logarithms. The growth-factor of the curve from which the table is constructed is called the "base" of the logarithms.

This method of constructing a table of logarithms and the uses of the table when constructed are explained in ch. XXXII., B, and Ex. LIV, B. The uses of a table of antilogarithms are illustrated at the same time. This table may be regarded as giving the ordinates corresponding to an arithmetic progression of abscissæ, just as the former table gave the abscissæ corresponding to an arithmetical progression of ordinates.

Some excellent teachers deprecate the introduction of the antilogarithmic table and give reasons against it which certainly carry weight. The author favours its retention less on practical grounds than because of its usefulness in developing the theory of the logarithmic and exponential functions. This usefulness appears very clearly in the discussion of ch. XXXIV.

Ch. XXXIII., A, and Ex. LV carry the theory of the subject an important step forward, for they show how the notion of logarithms can be freed from its dependence upon the growth curve and become purely arithmetical. We have, that is, now reached the concept of the logarithm of a number N as it presented itself to Napier—namely, as a measure of the "number of ratios" needed to bring us from 1 to N along a certain geometric series. As an example suppose that 1·1 is chosen to be the common ratio of the series of numbers. Then if underneath each term we write the "number of ratios"

from the beginning of the series up to that term this number is, by definition, the logarithm of the term above it:—

1	1·1	1·21	1·331	1·464	1·611	1·772	1·949	2·144	2·358	2·594
0	1	2	3	4	5	6	7	8	9	10

It is evident that upon this plan all logarithms must be integers and that the difference between one table of logarithms and another will consist simply in the fact that different ratios are used in constructing the series. But without changing the value of this ratio it is possible, by a slight complication of the original idea, to vary the logarithms at will. The complication consists in supposing the series 0, 1, 2, 3, . . . to measure not only the "number of ratios" in the geometric series but also the *number of steps* in an auxiliary arithmetic series which may be supposed to be written underneath it, term by term. If the name "logarithm" is—with some violence to its original meaning—transferred to the terms of the arithmetic series it becomes evident that any term in one and the same geometric series may have an endless number of logarithms varying with the value of the number taken as the common difference of the arithmetic series. Thus if 0·1 is taken as the common difference the logarithm of 1·611 will be 0·5 and that of 2·594 will be 1; if the difference is 0·2 the logarithms of the same two numbers will be respectively 1 and 2; with a difference of 0·3 the logarithms become 1·5 and 3; and so on endlessly.

So long as we consider only a few terms of the geometric series the subsidiary arithmetic series appears to be an unnecessary complication. But if we consider an actual table of logarithms suitable for serious use its value becomes evident. To be of practical value the table of logarithms must be constructed by means of a ratio very near to unity, so that no number which can be used in a calculation lies far from one of the terms of the geometric series. Ideally the terms of the latter should be so close that, for practical purposes, any given number may be considered identical with one of them. But in this case, or in any case that approximates to it, the "number of ratios" would soon become uncomfortably large. It is convenient, therefore, to divide all the "numbers of ratios" by some large number—preferably a high power of 10—so as to reduce them to an arithmetic series of decimals of which any suitable number of places may be retained

and the rest ignored. The most convenient way of describing the relation between the geometric series and the arithmetic series is to specify the value of the term of the former which corresponds to the term 1 in the latter. This number is called the "base" of the logarithms. In the first two of the three cases considered above the bases are, in order, 2·594 and 1·611. In the third case the base does not appear directly in the table. Since, however, the logarithm of 2·594 is 3 it is evident that 1 must be the logarithm of the cube root of 2·594 —that is, of 1·374. This number is, then, the base.

§ 5. *Napier's Logarithms.*—In substance the foregoing article is an exposition of Napier's theory of logarithms. There is, however, an interesting subtlety in Napier's argument which is omitted because it tends to blur the clear notion of the "number of ratios" as Napier seems at first to have conceived it. The reader will remember that in Napier's scheme the two points start moving from B and L in fig. 77 with the same velocity, and that if the points marked a are those reached at the end of the first unit interval of time then La is the logarithm of Aa. Now since the speed of the point in BA is constantly falling off La must be greater than Ba. On the other hand, if B′ be supposed to mark the position of the upper point a unit interval before it reached B, then La must for the same reason be less than B′B. Let r be the common ratio of the lines or sines AB, Aa, Ab, etc. Then, since by hypothesis AB is unity, $Aa = r$ and $Ba = 1 - r$. Also $AB' = AB/r = 1/r$, so that $B'B = 1/r - 1 = (1 - r)/r$. Thus the logarithm L$a$ lies between $1 - r$ and $(1 - r)/r$. In Napier's calculations r was taken to be 0·9999999, in which case La lies between 0·0000001 and the same number divided by 0·9999999, or 0·00000010000001. There will, therefore, be little error in taking it to be the mean of these numbers or 0·000000100000005. This last number is adopted as the common difference of the arithmetic series of logarithms while 0·9999999 is taken as the common ratio of the geometric series of numbers or antilogarithms "adapted" to it. Thus the logarithms of 0·99999980000001 and 0·99999970000003, which are the next two numbers of the geometric series, must be respectively 0·00000020000001 and 0·000000300000015.

The advantages of this procedure over the simpler one described above are that it assigns theoretically a logarithm to every number and indicates a method of calculating it

within limits which may be made as small as we please. According to the scheme of § 4 the only logarithms given directly are those of the members of the geometrical series ; the logarithm of another number is given only if, for practical purposes, it may be considered as coinciding with a term of that series. But Napier's concept of the two moving points obviously assigns a position for the " logarithm point " on the line L corresponding to every possible position of the " number point " on the line AB. Again Napier's method fixes the base of the logarithms upon a definite principle. Suppose, for instance, that in accordance with the method of § 4 the ratio is taken to be 0·9 and the common difference of the logarithms 0·1. Then the base will be $(0{\cdot}9)^{10}$ or $(1 - \frac{1}{10})^{10}$. If a finer scale of numbers is to be secured we may take 0·99 as the ratio and 0·01 as the difference. The base will in this case be $(1 - \frac{1}{100})^{100}$. Adopting the same general plan but aiming at still closer geometric scales we may take 0·999, 0·9999, etc., as the ratio and 0·001, 0·0001, etc., as the common difference. Thus the successive bases will be $(1 - \frac{1}{1000})^{1000}$, $(1 - \frac{1}{10000})^{10000}$, etc. These numbers have as their " limit " the number which mathematicians denote by the symbol e^{-1} and which = 1/2·7182818 . . . ; but Napier's base, since it corresponds to a geometrical scale constructed upon the same plan but of infinite closeness, is the limit itself. It is easy to see that this consequence follows from the hypothesis that the two points of fig. 77 start their movements with the same velocity. For if the point a be taken on AB so that $Aa = 1 - (1/10)^n$ then La is nearer to $(1/10)^n$ the greater the value of n ; complete equality between Ba and La being the limit approached as n approaches infinity.

§ 6. *Common Logarithms* (Exs. LVI, LVII ; ch. xxxiii., B, C.).—The next two exercises are devoted to illustrating the special convenience of logarithms to base 10 and to giving practice in the use of tables of such logarithms. The treatment given in ch. xxxiii. is too full to need further elucidation, but the teacher may be recommended to follow rather carefully the methods suggested for determining the characteristic of a logarithm or the number of figures before the decimal point in an antilogarithm. After a little practice these methods may, no doubt, be abbreviated to the more familiar rules usually prefixed to books of mathematical tables ; but these rules will be used with more intelligence if the students go

through a period of drill in which they are required to give full analyses of their procedure. From this point of view it is much to be desired that in tables of logarithms the "argument," and in tables of antilogarithms the logarithms, should always be printed as a number in the "standard form" (see p. 175). This useful practice is adopted in the logarithms and antilogarithms published as a separate volume of this series. It is also a feature of the excellent American tables of Prof. E. V. Huntingdon.

§ 7. *The Logarithmic and Antilogarithmic Functions. Fractional Indices* (Ex. LVIII; ch. xxxiv.).—The last few exercises have been concerned entirely with the practical aspect of logarithms. We now turn to the theoretical aspect and study the connexion between a number and its logarithm as a function possessed, like the parabolic or the hyperbolic function, of definite and distinctive properties. In connexion with the argument of ch. xxxiv., § 2, the teacher should consult ch. xli., p. 432.

Just as the practical problem of finding the logarithm for any given number suggests the idea of the logarithmic function of the variable x, so the inverse problem of finding the number corresponding to any given logarithm leads to the notion of the antilogarithmic function. This part of the argument of ch. xxxiv. is of great importance. Its crux is the question of the symbolic representation of the antilogarithmic function. The first obvious mode of representation is

$$y = \text{antilog}_a x$$

and this would always be sufficient, though cumbersome. But an examination of the properties of the function shows that they are identical for *all* values of x with the properties which the function a^x possesses for *integral* values of x. The discovery of this fact has its natural sequel in the proposal to extend the range of application of the symbolism a^x to include fractional values of the variable. The consequences of the extension are sufficiently described in § 4.

Division A of Ex. LVIII is intended to illustrate these theoretical questions. Division B consists of examples which bring out the usefulness of the index-notation for the antilogarithmic function. The method of investigating certain physical laws illustrated by **Nos. 16-21** is of considerable interest and practical importance, especially to engineers.

It is fully described by Prof. Perry in his *Practical Mathematics* (Lecture IV).

Nos. 22-27 may be regarded as illustrating in a simple way the method by which Henry Briggs first calculated logarithms to a given base (see the note, *Exercises*, p. 296). Mr. Edwin Edser appears to have been the first person to propose the method of graphic interpolation of No. 26.

The exercise ends with a few simple examples on the manipulation of fractional indices—a subject to which too much attention has been given in text-books and examinations.

§ 8. *The number "e"* (Ex. LIX; ch. xxxv.).—The ideas connected with the symbol "e" are of such importance in algebra that they should certainly find a place, if possible, even in the elementary course. The general argument of Section III makes it a comparatively easy matter to deal with them in a way which brings out their real significance and yet makes very moderate demands upon the student. Abstracting from the concrete setting of the argument of ch. xxxv. we see that it leads to the important conclusion that as n increases the value of $(1 + i/n)^n$ grows constantly nearer to a definite number which it never actually reaches, though it comes and ever afterwards remains nearer to it than any number that can be named, however small that may be. In the language to be used at a later stage of the work, $(1 + i/n)^n$ is shown to have a definite "limit". It is also shown that for different values of i that limit can be expressed as e^i where e is the number related to $(1 + 1/n)^n$ in the same way as the former number is related to $(1 + i/n)^n$. It will be noted that the function of the geometrical argument is in the first place to give concrete significance to the expression $(1 + i/n)^n$ and in the second place to prove that it "tends to a limit". When this fact has been proved it is sufficient to obtain an approximation to e by the method suggested in No. 22.

The objection that geometrical reasoning should not be used in order to arrive at a conclusion in algebra will hardly be thought serious in the present connexion, and at the stage of the student's progress here in view. A simple and effective argument need not be rejected upon the academic ground that it mixes geometrical and arithmetical reasoning. It is, however, important that the assumptions upon which its cogency rests should be clearly understood, and for this

reason the teacher should not fail to point out that the argument assumes without proof that a "growth-curve" has at each point a definite tangent. At a later stage the desire of eliminating this assumption may be made the motive of an attempt to discover a purely algebraic analysis. At the present stage such an attempt would find little support in any logical scruples on the part of the student.

CHAPTER XXX.

THE GRAPHIC SOLUTION OF GROWTH PROBLEMS.

(Ch. XXIX., § 2; Ex. LII.)

§ 1. *Two Measures of Magnitude-Change.*—The problems of Ex. LI suggest that there are two distinct ways of measuring the change in magnitude which a constantly increasing or decreasing thing undergoes during a given interval. Let the magnitudes at the beginning and the end of the interval be m_1 and m_2. Then $m_2 - m_1$ gives one measure of the change while m_2/m_1 gives the other. The first may be called the *growth-difference*, the second the *growth-factor*. If the thing has increased during the interval, the growth-difference will be positive and the growth-factor greater than one. If it has decreased, the growth-difference will be negative and the growth-factor less than one.

Of these two measures of change sometimes the one, sometimes the other, is the more important. Imagine, for example, a vessel into which water is running. Let the amount in the vessel at the beginning be 10 gallons and at the end of 1, 2, 3, 4, etc., minutes subsequently be 12, 14, 16, 18, etc., gallons. Then the growth-differences for all these intervals are the same, namely 2 gallons, but the growth-factors are different, namely 1·20, 1·17, 1·14, 1·125, etc., approximately. It is evident that in this case the statement about the growth-differences is more useful and interesting than the statement about the growth-factors. On the other hand, consider the heights of the average girl between the ages of $8\frac{1}{2}$ and $14\frac{1}{2}$ (*Exercises*, p. 269). The successive growth-differences in this case are 2·0, 2·0, 2·1, 2·3, 2·4, and 2·5 inches, while, as you saw in Ex. LI, No. 5, the growth-factor has practically the same value for each of these years—namely, 1·042. In this case, then, the growth-factor is the more important measure of change.

§ 2. *Two Laws of Continuous Magnitude-Change.*—What we have said about these examples would be equally true if the increase measured took place in "jerks" between the measurements. As a matter of fact, however, both the filling of a vessel and the growth of a girl are processes of *continuous* increase. From the facts that in the former case the growth-differences, and in the latter the growth-factors, are constant for measurements separated by a certain constant interval, we cannot, of course, infer that the same regularities would be exhibited if another interval were chosen. Nevertheless, the examples *suggest* two types of continuous increase in which these regularities would hold good respectively for *all* equal intervals, however large or small. Let a series of measurements of a growing thing be made at moments separated by equal intervals of time. Then if, no matter what interval is chosen, each measurement is greater (or less) than its predecessor by a constant amount the growth is an instance of the former type; if the ratio of each measurement to its predecessor is constant it is an instance of the second type.[1]

§ 3. *Problems of the Second Type.*—Suppose that the filling of the vessel in § 1 followed the first type of growth. Let it be given that at a certain moment t_0 the vessel contained 20 gallons and that the growth-difference was 2 gallons/min. Then the quantity of water at any other moment can, of course, easily be calculated. For instance, $3\frac{1}{2}$ minutes after t_0 the quantity will be $20 + 2 \times 3\frac{1}{2}$, $5\frac{1}{4}$ minutes before t_0 it must have been $20 - 2 \times 5\frac{1}{4}$. In making these calculations we argue that the growth-difference for half a minute must be one-half of 2 gallons, for a quarter of a minute one-quarter of 2 gallons, etc. It is obvious that any other problem involving this kind of growth could be solved by a similar method.

Next suppose, on the other hand, that the growth of "the average girl" is of the second type. Let it be given that at $11\frac{1}{2}$ her height is 53·8 inches and that the yearly growth-factor is 1·042. How shall we calculate her height, say at 14 years?

The height at $13\frac{1}{2}$ (i.e. after two years) is, by hypothesis, given by the calculation

$$53{\cdot}8 \times 1{\cdot}042 \times 1{\cdot}042 = 58{\cdot}4 \text{ inches.}$$

[1] It may, of course, belong to neither type.

The difficulty is to find how to deal with the extra half-year. It will not do, of course, to find how much the girl will grow during the next year and add one-half of this amount to the above product. To act so would be to assume that during this year her growth belongs to the *first* type—which is contrary to our supposition. Nor can we argue that the growth-factor for six months is one half of the yearly factor. This argument would lead us to find the girl's height at 14 by multiplying her height at $13\frac{1}{2}$ by 0·521—that is, we should conclude that in these six months her height is *reduced* by nearly a half! Ex. LI, Nos. 15-21, suggest the correct method to follow. Let the growth-factor *for six months* be called r. Then we have

$$\text{height at } 14 = 58{\cdot}4 \times r$$
$$\text{height at } 14\tfrac{1}{2} = 58{\cdot}4 \times r \times r$$

But, by hypothesis,

$$\text{height at } 14\tfrac{1}{2} = 58{\cdot}4 \times 1{\cdot}042$$
$$\text{hence } r^2 = 1{\cdot}042$$
$$\text{and } r = \sqrt{1{\cdot}042}$$

By a similar argument we could calculate the height, say at $10\frac{1}{4}$. The height at $10\frac{1}{2}$ is 53·8/1·042 = 51·6 inches. This number must be divided by r, the growth-factor (in this case) for a quarter of a year. To find r we note that four successive divisions by r should give us the height at $9\frac{1}{2}$. But this height could also be obtained by a single division by 1·042. Hence $r^4 = 1{\cdot}042$ or $r = \sqrt[4]{1{\cdot}042}$. The general rule is easily derived from these examples. The growth-factor for $\frac{1}{n}$ of a year is always the nth root of the growth-factor for a year.

The foregoing rule can be applied without much difficulty in simple cases but seems of little use in others. For instance, how are we to find the growth-factor for $\frac{40}{73}$ of a year, so as to calculate the "average girl's" height 200 days after the last measurement? Even to calculate the growth-factor for a month (taken as $\frac{1}{12}$ of a year) we must go through the tedious operation of finding the square root of the square root of the cube root of 1·042.

§ 4. *Graphic Solutions: the "Growth-Curve".*—To avoid these difficulties we naturally fall back upon the graphic method. This method could, if it were worth while, be used to solve problems of the first type. Given one magnitude of

the growing thing and the growth-difference for a certain interval we could at once draw the straight line whose points represent all other magnitudes of the thing. In problems of the second type the graph will not, of course, be a straight line. But if we plot a number of points, representing the magnitudes calculated by means of the growth-factor for a given interval, we may reasonably expect that the smooth curve through these points will give us the magnitudes which it is tiresome or impossible to determine by calculation.

The results of Ex. LI, No. 14 (ii), set out in the following table, may conveniently be utilized in testing this expectation. Imagine a continuously increasing quantity subject to a constant growth-factor of 1·25 for the unit of time, and let its present magnitude be unity. The second line of the table gives the magnitude of the quantity 1, 2, . . . 5 units of time ago and 1, 2, . . . 5 units hence. The third line of figures gives, in a similar way, the history of a continuously decreasing quantity, subject to a constant growth-factor of 0·8 for the unit of time. Since multiplication by 0·8 is the same as division by 1·25 and *vice versa*, one of these rows is simply the other row reversed.

n	− 5	− 4	− 3	− 2	− 1	0	+ 1	+ 2	+ 3	+ 4	+ 5
$1 \times (1{\cdot}25)^n$	0·328	0·410	0·512	0·640	0·800	1	1·250	1·563	1·953	2·441	3·052
$1 \times (0{\cdot}8)^n$	3·052	2·441	1·953	1·563	1·250	1	0·800	0·640	0·512	0·410	0·328

The members of the class should undertake the plotting of the curves in pairs. The first member of each pair should plot the successive magnitudes given by a growth-factor of 1·25, and his companion those due to a growth-factor of 0·8. Each should then draw a smooth curve through his points with the greatest possible care. The curves should, of course, be identical, except that corresponding parts lie on opposite sides of the y-axis.

The hypothesis which we have now to test is that these curves are graphic expressions of the law exhibited in continuous growth of the second type. If this hypothesis is correct the ratio between the heights of equidistant ordinates should be the same in all parts of the curve, for these ratios

measure the growth-factors of the quantity during equal intervals of time. Each pupil should draw in different parts of his curve three pairs of equidistant ordinates, the distance between them being selected at random. He should determine the ratios obtained by dividing the height of the right-hand member of each pair of ordinates by that of the left-hand member. If his curve is drawn finely and accurately he will be able to satisfy himself that the three ratios are equal. Since similar results are obtained, with different sets of ordinates, by all the class, there can be no reasonable doubt that the curves really do give the magnitude of the changing quantity at all times within the scope of the graph.

§ 5. *The Growth-factor necessarily Positive.*—We may for convenience refer to these graphs as "growth-curves," and to the kind of continuous change which they represent as "uniform growth with constant growth-factor". This expression will distinguish growth of the second type from the "uniform growth with constant growth-difference" which constitutes the first type.

The graphs help to bring out certain important differences between the two types. In the first type there is no lower or upper limit to the magnitude of the quantity, for the graph is a sloping straight line. In the second type, on the other hand, while the magnitude may increase without limit it cannot decrease without limit. A large number of divisions by a factor greater than 1 or multiplications by a factor less than 1 will reduce its measure below any given *positive* number, but can never make it zero or negative. The x-axis is, therefore, asymptotic to the growth-curve.

If the original magnitude is positive the whole of the growth-curve will be above the x-axis; if negative, below. It is important to notice that the growth-factor must itself always be positive in these problems. Thus, let a unit magnitude be subject, if possible, to a growth-factor of -2. Then its magnitudes at the end of successive units of time will be: $1 \times (-2)$, $1 \times (-2)^2$, $1 \times (-2)^3$, etc., i.e. -2, $+4$, -8, $+16$, etc. The corresponding past magnitudes will be $-\frac{1}{2}$, $+\frac{1}{4}$, $-\frac{1}{8}$, $+\frac{1}{16}$, etc. Now it is, of course, conceivable that a changing quantity should have these magnitudes at the end of successive equal intervals, but its "growth" would be very different from the constantly increasing or decreasing magnitude which we have been studying. It would be a "growth" which now

makes the thing larger and then smaller and then larger again, etc., like the movement of the water up and down a sea-wall. We shall find it convenient to ignore such kinds of "growth" for the present [1] and to confine our attention to cases where the growth-factor is positive.

[Ex. LII should now be worked.]

[1] They will be considered in *Exercises*, Part II, Sections VI, VII.

CHAPTER XXXI.

THE GUNTER SCALE.

(Ch. xxix., § 3; Ex. LIII.)

§ 1. *A Multiplicity of Curves unnecessary.*—Problems in Ex. LII which involved different growth-factors were solved by means of different curves. When the growth-factor was 1·3 curve A of *Exercises*, fig. 50, was used, when it was 1·25 reference was made to curve B, and so on. If it were really necessary to have a new growth-curve for every problem which involved a new growth-factor a large collection of such curves would be required. Fortunately it can be shown that no such necessity exists. For example, suppose that we had only curve C (in which the growth-factor is 1·1) and that we wanted to solve problems involving a growth-factor of 1·3. Inspection shows that an ordinate of height 1·3 is to be found in this curve where the abscissa is 1·88. It follows, from the fundamental property of growth-curves, that whatever pair of ordinates is taken at a distance from one another of 1·88, the ratio of the longer to the shorter will in every case be 1·3. Thus the height of the ordinate at + 3·76 is 1·69 (i.e. the height of the ordinate in curve A whose abscissa is + 2), the ordinate at − 1·88 is 0·77 (i.e. that of the ordinate in curve A whose abscissa is − 1). It is possible, then, to determine the magnitude of an original unit after it has been increasing or decreasing for a time t with a growth-factor of 1·3 by means of curve C as well as by curve A. The required magnitude—which is the height of the ordinate whose abscissa is t in curve A—will also be the height of the ordinate whose abscissa is $1{\cdot}88t$ in curve C.

Again, suppose that we had only curve A and wished to solve a problem in which the growth-factor is 1·2. The ordinate which has this value is distant 0·61 from the origin. The magnitude of the original unit after time t will, therefore,

be given by the ordinate whose abscissa is $0{\cdot}61t$. For example, after $3\frac{1}{2}$ unit intervals its magnitude will be that of the ordinate whose abscissa is $(+\ 0{\cdot}61) \times (+\ 3\frac{1}{2}) = +\ 2{\cdot}13$, i.e. 1·89. Similarly its magnitude $7\frac{1}{3}$ intervals ago was that of the ordinate whose abscissa is $(+\ 0{\cdot}61) \times (-\ 7\frac{1}{3}) = -\ 4{\cdot}27$, i.e. 0·38. Generalizing, we see that a problem involving *any* growth-factor r, can be solved by means of *any* growth-curve. Let p be the abscissa of the point whose ordinate is r. Then the magnitude of an original unit at time t is the height of the ordinate whose abscissa is pt.

§ 2. *The Gunter Scale.*—The preceding argument suggests that it may be profitable instead of drawing a different growth-curve for each growth-factor, to draw *one* such curve with exceptional care and to use it for all growth-problems. Curve A may be selected for this purpose on the ground that, since it rises more rapidly than the others, a given length of base offers a wider range of ordinates. Supposing it is to be adopted as the working curve it is worth while to facilitate its use by a simple device. This consists in recording at suitable points along the horizontal axis the heights of the ordinates at those points. By this means the abscissa corresponding to a given ordinate can be found much more rapidly than by consulting the vertical scale.

The method is illustrated in *Exercises*, fig. 50. In order to obtain a still wider range of ordinates the vertical axis is supposed to be moved to the position of the dotted line, and the vertical scale to be contracted.[1] The graduations at different points of the line GG simply record, upon the new scale, the heights of the ordinates immediately above them. Thus the graduation "2" is directly below the ordinate whose height is 2 in the contracted scale, the graduation "7·3" directly below the ordinate whose height is 7·3, and so on.

In graduating the line GG no notice has been taken of the values of the abscissæ of the various ordinates. It is, in fact unnecessary to record them. Suppose, for example, that we have a problem in which the growth-factor is 1·42. Applying a centimetre rule to GG you will find that the ordinate whose height is 1·42 is situated exactly 2 cms. from the origin of the scale. The magnitude of an original unit after 2, 3, 4 . . . , time-intervals will therefore be given by the gradua-

[1] If the size of the page had been unlimited it would, of course, have been simpler and better to continue the curve to the right.

tions situated 4 cms., 6 cms., 8 cms., . . . along the line; for these graduations give the ordinates whose distance from the dotted axis of the curve is 2, 3, 4 . . . times the distance of the ordinate 1·42. Similarly, the magnitude of the unit after 1·2 intervals, 2·6 intervals, 3·05 intervals, etc., is given by the graduations 2·4 cms., 5·2 cms., 6·1 cms., etc., from the origin of the scale.

It is obvious from these examples that the scale might be cut out and used without further reference to the curve from which it has been constructed.

The first man to graduate a line in this way and to use it for calculations was Edmund Gunter (1581-1626), an English mathematician who invented many improvements in the arts of navigation and surveying. Among these was the surveyor's chain, still called "Gunter's chain". A line graduated like GG of fig. 50 is the essential part of a calculating apparatus which has been much used by sailors and is called by them "Gunter's scale".

§ 3. *Extensions of the Scale.*—The graduations of the Gunter scale in *Exercises*, fig. 50, range only from 1 to 10. It is, however, extremely easy to extend them both above 10 and below 1.

Imagine the scale on the dotted axis of Curve A to be contracted 10 times—that is, let the figures 1, 2, 3, . . . be replaced by the figures 10, 20, 30 . . . It is obvious that each graduation on the Gunter scale must now be multiplied by 10 in order to give the length of the ordinate above it. Similarly, if the vertical scale be contracted 100 times each graduation on the Gunter scale must be multiplied by 100. Thus to construct a Gunter scale graduated from 1 to 1000, a line must be taken three times as long as GG; the mode of division of the first or fundamental section of this line must be repeated in each of the other two sections, but the graduations in the second section must be 10 times, and those in the third section 100 times as high as the corresponding graduations in the first section. It is obvious that the only reason why the graduations could not thus be continued indefinitely is that, since the scale in each section is ten times as contracted as in the preceding section, the subdivisions would soon become too close together to be read with accuracy.

In order to obtain the graduations below 1 the scale on the dotted vertical axis may be supposed to be *expanded* 10 times.

The measure of each ordinate recorded on GG will thus be reduced 10 times, so that the graduations will run from 0·1 to 1. If the vertical scale be expanded 100 times the graduations will be reduced to the range from 0·01 to 0·1, and so on. Thus the Gunter line may be continued to the left of the fundamental section in sections whose length is the same as that of GG, the scale of the graduations in each section being 10 times as expanded as that in section to the right of it. Theoretically, then, the graduation can be extended to the left without limit. Practically, the subdivisions would soon become too far apart to be used with convenience.

§ 4. *The Gunter Scale in Growth-problems.*—It was seen in § 2 that a Gunter scale can be used, in conjunction with a centimetre or inch rule, to solve any growth-problem. The operations can be made extremely simple by the method illustrated in the following examples.

Example 1.—To find the magnitude of an original unit after 5·2 years, the annual growth-factor being 1·4.

Take a strip of paper, lay it along the Gunter scale with one end on the graduation "1," and mark on the edge the point where the graduation "1·4" falls. Now lay the strip across a sheet of squared paper, as shown in fig. 79.[1] Place the end of the strip anywhere on the vertical marked "0" and swing it round until the mark on the edge lies upon the vertical through the graduation on the squared paper marked "1". Mark on the edge of the strip the point P, where it is crossed by the vertical whose graduation is 5·2. Once more lay the strip along the Gunter scale. The reading against the mark P is the magnitude required. It will be found to be 5·75.

Example 2.—An original unit increases to a magnitude of 33·6 in 7·5 units of time. What is the growth-factor?

By § 3 the graduation "33·6" occurs in the second section of the Gunter scale in the position occupied by 3·36 in the fundamental section. It is necessary, therefore, to mark along the edge of the strip a distance equal to the whole length of GG *plus* the length from the beginning to the graduation "3·36". Lay the strip across the squared paper as before and swing it round until the mark on the edge lies upon the vertical whose graduation is "7·5". Mark the point where the edge crosses the vertical graduated "1". Once more lay

[1] For clearness only the unit lines of the squared paper are shown in the figure.

the strip along the Gunter scale. The reading against the second mark gives the growth-factor. It should be 1·6.

Example 3.—The magnitude of a quantity decreases in 4·8 years from unity to 0·46. Find the growth-factor.

We are here concerned with the first section of the Gunter scale to the left of the graduation "1". In this section the graduation "0·46" would occupy the same position as "4·6"

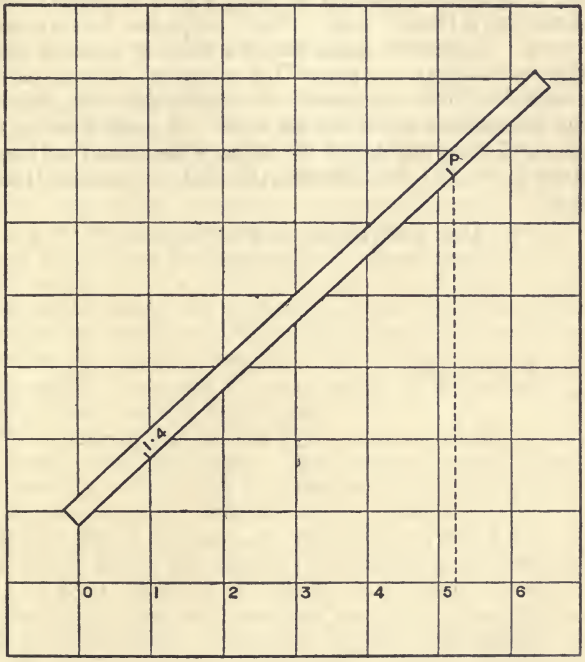


FIG. 79.

occupies in GG while the graduation "1" would occupy the position of the present graduation "10". The beginning of the strip must, therefore, be placed against the graduation "10" in GG, and the point on the edge marked which lies against the graduation "4·6". As before, the strip must be placed with its beginning on the zero graduation of the squared paper and must be swung round until the mark on the edge

lies on the vertical whose graduation is "4·8". The edge must be marked where it crosses the vertical graduated "1". When the strip is again laid on the Gunter scale from right to left (i.e. with its beginning on the graduation "10") the second mark lies against the graduation "8·5". The growth-factor must, therefore, be 0·85.

The risk of error involved in marking the auxiliary strip could, of course, be avoided by using a strip of paper already graduated as a Gunter scale. Such strips can be very easily procured. Engineers make use of a kind of squared paper (called "semi-logarithm paper") in which the lines of one set of parallels follow one another in exactly the same positions as the graduations of the Gunter scale. A single sheet of this paper can be cut into about 20 strips which need only to be graduated like GG in *Exercises*, fig. 50, to become Gunter scales.

[Ex. LIII should now be worked.]

CHAPTER XXXII.

LOGARITHMS.

A. The Slide Rule (ch. xxix., § 3; Ex. LIV, A).

§ 1. *Multiplication and Division by Means of Growth-curves.*—The curves of *Exercises*, fig. 50 (or the Gunter line derived from them) were used in solving the problems of Ex. LIII as the only really practicable means of solution available. It is now to be shown that the property of growth-curves which made them indispensable in these calculations can be used to lessen in a remarkable way the labour of certain calculations which are generally carried out by the ordinary processes of arithmetic. These further applications all depend upon the fact that the heights of two ordinates at a given distance apart, no matter in what region of the curve they are taken, always have a fixed ratio.

This property can, in the first place, be used to facilitate processes of multiplication. As an example consider how the product $1{\cdot}9 \times 1{\cdot}3$ could be obtained by the aid of curve A (*Exercises*, fig. 50). The first thing to do is to pick out the ordinates whose heights are respectively 1·9 and 1·3. They are found where the abscissæ are respectively 2·45 and 1. But, by the property of the curve, an ordinate situated 1 unit to the right of any given ordinate will always be 1·3 times as high as that ordinate. Thus the height of the ordinate whose abscissa is 3·45 must be $1{\cdot}9 \times 1{\cdot}3$. But the height of that ordinate is 2·47. It follows, that $1{\cdot}9 \times 1{\cdot}3 = 2{\cdot}47$.

The same problem could have been solved by means of curve B. The ordinates of height 1·9 and 1·3 have respectively 2·9 and 1·2 for their abscissæ. In this case, then, an ordinate 1·3 times as high as a given ordinate will always be found at a distance of 1·2 to the right of that ordinate. Thus the product $1{\cdot}9 \times 1{\cdot}3$ is given by the height of the ordinate whose abscissa is $2{\cdot}9 + 1{\cdot}2 = 4{\cdot}1$. The height of this

ordinate is 2·48. The values for the product obtained from the two curves differ by less than $\frac{1}{2}$ per cent.

In general, then, we have the following method of determining by means of *any* growth-curve the product of two numbers P and Q. Find the abscissæ, p and q, of the ordinates whose heights are respectively P and Q. Then PQ is the height of the ordinate whose abscissa is $p + q$.

The result of a division process can obviously be obtained by a similar method. The height of the ordinate situated q to the *left* of a given ordinate will always be equal to the height of the latter *divided* by Q. Hence the ordinate whose height is P/Q is the ordinate whose abscissa is $p - q$. For example let P = 1·3 and Q = 1·9. Then in curve A, $p = 1$, $q = 2·45$ and $p - q = -1·45$. The ordinate whose abscissa is $-1·45$ gives, then, the value of the quotient. It will be seen to be 0·68. In curve B, $p = 1·2$, $q = 2·9$ and $p - q = -1·72$. The ordinate with this abscissa has, again, a height of 0·68.

§ 2. *Multiplication and Division by the Gunter Scale.*—As might be supposed, products and quotients can be obtained still more readily by means of the Gunter scale. The validity of the following rules is obvious from ch. xxxi.

To find P × Q mark off on a strip of paper the distance from the beginning of the scale to the graduation "Q". Transfer the beginning of the strip to the graduation "P". Then the mark on the edge lies against the graduation which measures PQ. To find P/Q place the mark on the edge against the graduation "P," then the beginning of the strip lies against the graduation which measures P/Q.

In two cases difficulty will arise. (*a*) The product PQ may be greater than 10 so that the mark on the strip lies beyond the scale on the right. The mark must in this case be supposed to lie in the first section to the right of the fundamental section and the graduation which would be against it must be determined by the principle of ch. xxxi., § 3. Mark on the strip the point against the graduation "10". Carry this point back to the graduation "1". Then the former mark on the strip lies against a graduation which is simply ten times as small as the corresponding graduation in the second section. Thus, if it lies against the graduation "2·35," the product in question is 23·5.

An alternative way to overcome this difficulty is as follows. Cut a strip of paper exactly as long as GG and, as before,

mark on it the position of the graduation " Q ". Reverse the strip and apply it to the Gunter line so that what was the beginning of the strip coincides with the graduation " 10," while what was originally its right-hand end coincides with the graduation " 1 ". Now slide the strip to the right (fig. 80) until the " Q " mark again coincides with the graduation " P ". The present right-hand end of the strip now lies against what would be the graduation " PQ " in the first section to the right of GG if that section were there. But since the mark at the (present) left-hand end of the strip has moved to the right through the same distance as the other end it coincides with the graduation in the fundamental section of the Gunter scale which corresponds with the graduation " PQ " in the (hypothetical) section to the right of it. The value of PQ is, therefore, the value of this graduation multiplied by 10.

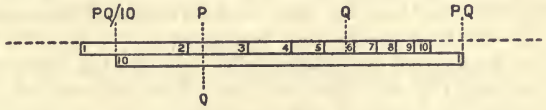


FIG. 80.

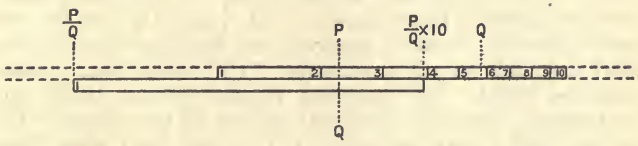


FIG. 81.

(*b*) The second difficulty arises in determining P/Q where Q is greater than P. In this case, when the mark on the strip is made to coincide with the graduation " P," the beginning of the strip, which should now coincide with the graduation " P/Q," lies outside GG (fig. 81) in the (hypothetical) section of the Gunter scale to the left of the graduation " 1 ". Its position can, however, be determined by an argument similar to the one employed above. The right-hand end of the strip lies just as far to the left of the graduation " 10 " as the beginning of the strip lies to the left of the graduation " 1 ". It coincides, therefore, with a graduation which is exactly ten times as great as P/Q. For instance, if that graduation is 8·7, P/Q = 0·87.

§ 3. *The Slide Rule.*—A very obvious improvement on these methods is to employ, instead of the strip of paper, a second Gunter line graduated on the same scale as the first. A pair of such scales, arranged so that one can slide backwards and forwards beside the other, constitutes a Slide Rule—a calculating apparatus which is constantly used by engineers, architects, etc. The easiest way to make a slide rule is to cut two strips of "semi-logarithm paper," to graduate both like GG in *Exercises*, fig. 50, and to paste them on to two strips of cardboard or wood. One of these strips must be fastened down to a cardboard or wooden base which must be wide enough to accommodate also the second strip. (For temporary use it is sufficient to pin the fixed scale to a drawing-board or the desk; even the strips of cardboard or wood can be dispensed with.)

The product and quotient of two numbers P and Q, each of which is less than 10, can be determined precisely as in § 2. To obtain a product the beginning of the movable scale must be set against the graduation "P" on the fixed scale. The product PQ will then be the graduation on the fixed scale which lies against the graduation "Q" on the sliding scale. If the product is greater than 10, the sliding scale must be reversed, the graduations "P" and "Q" on the two scales must be brought together, and the graduation of the fixed scale against the graduation "10" of the sliding scale must be noted. The required product is ten times this graduation. To obtain the quotient P/Q set the reading "Q" on the sliding scale against "P" on the fixed scale, both scales being held so that their graduations increase to the right. If Q is less than P the quotient is the graduation of the fixed scale which lies against the graduation "1" on the sliding scale. If Q is greater than P the quotient is ten times *less* than the graduation against the "10" of the sliding scale.

If either P or Q is greater than 10 it must first be expressed in the "standard form". The product or quotient of the numbers less than 10 is then obtained and is afterwards multiplied (or divided) by the appropriate power of 10. Thus the quotient $7850/57 = (7{\cdot}85 \times 10^3)/(5{\cdot}7 \times 10) = (7{\cdot}85/5{\cdot}7) \times 10^2$. The quotient $7{\cdot}85/5{\cdot}7$ is first obtained (it is 1·38) and is then multiplied by 100 to give the (approximate) quotient 138.

[Ex. LIV, A, should now be worked.]

B. Logarithms (ch. XXIX., § 4 ; Ex. LIV, B).

§ 1. *Tables of "Logarithms".*—Convenient as the slide rule is as a "ready-reckoner" its use is subject to obvious drawbacks. The accuracy obtainable depends upon the exactness and fineness of the graduations and upon the correctness with which they are read. Speedy and sure calculations can be made only after considerable practice. Moreover a strong and trustworthy slide rule is an expensive instrument. For these reasons alone it would be worth while to examine another way in which the properties of growth-curves can be utilized in facilitating calculations.

As we have seen, a Gunter scale is a graphic record of the heights and positions of the ordinates of some growth-curve. The "ready-reckoner" now to be studied is one in which the same facts are recorded, not graphically, but in the form of a table. Here is part of such a table, constructed from curve A, *Exercises*, fig. 50.

LOGARITHMS TO BASE 1·3.

n	log	*n*	log	*n*	log
1·0	0·00	1·6	1·79	2·2	3·00
1·1	0·36	1·7	2·00	2·3	3·17
1·2	0·69	1·8	2·24	2·4	3·33
1·3	1·00	1·9	2·44	2·5	3·49
1·4	1·28	2·0	2·64	2·6	3·64
1·5	1·55	2·1	2·83	2·7	3·78

In a table of this kind the "argument" or number which the calculator has in mind when he refers to it is the height of an ordinate of the growth-curve, while what he wants to find out from the table is the abscissa of this ordinate. The first column, therefore, headed "number" (*n*), contains the heights of ordinates while the second column gives the corresponding abscissæ. The abscissæ are here called *logarithms*.[1] It is obvious that the logarithm of unity is zero in all growth-curves, but that the logarithms of all other numbers will depend upon the growth-factor of the curve. For example

[1] This name was given to them by John Napier who invented these aids to calculation about 1594. He also called them "artificial numbers".

the logarithm of 1·5 is 1·55 in curve A, 1·86 in curve B and 4·28 in curve C. It is necessary, therefore, to specify the *base* or growth-factor employed in constructing the table. Thus the full description of the foregoing table will be: "A table of logarithms of numbers from 1 to 2·7 to the base 1·3".

§ 2. *The uses of Logarithms.*—A table of logarithms can be used for all the purposes for which the Gunter scale or slide rule can be employed.

Example 1.—Find the fifth root of 2·7. The logarithm or abscissa of the ordinate 2·7 is 3·78. By the property of the growth-curve the abscissa of the ordinate $\sqrt[5]{2·7}$ is one-fifth of this, or 0·756. This is not the logarithm of any number given in the table, but it lies between the logarithms of 1·2 and 1·3. Fig. 82 shows how its position between them can be calculated approximately. P_1N_1 and P_2N_2 represent the ordinates 1·2 and 1·3, and N_1N_2 the portion of the horizontal axis between the abscissæ 0·69 and 1·0. For convenience the abscissæ are represented on a larger scale than the ordinates. It is seen that the short length of curve between the ordinates differs little from the (dotted) straight line P_1P_2. Let pn be the ordinate of the straight line corresponding to the abscissa 0·756. Then its height is approximately the same as that of the ordinate to the curve which has the same abscissa. The height of pn is easily calculated by proportion :—

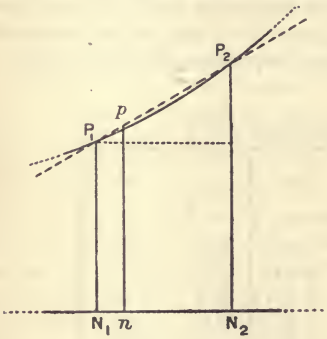


FIG. 82.

$$pn = 1·2 + 0·1 \times \frac{0·756 - 0·69}{1·00 - 0·69}$$
$$= 1·22$$

We conclude that 0·756 is approximately the logarithm of 1·22 so that $\sqrt[5]{2·7} = 1·22$ approximately.

This method of determining the number corresponding to a logarithm which lies between two logarithms given in the table is called "the method of proportional parts". It can,

of course, also be used to find the logarithm of a number which lies between two of the numbers given in the table.

Example 2.—Divide 2·425 by 1·733. By the method of proportional parts we have

$$\begin{aligned} \log 2{\cdot}425 &= 3{\cdot}33 + \tfrac{1}{4}\,(3{\cdot}49 - 3{\cdot}33) \\ &= 3{\cdot}37 \\ \log 1{\cdot}733 &= 2{\cdot}00 + \tfrac{1}{3}\,(2{\cdot}24 - 2{\cdot}00) \\ &= 2{\cdot}08 \end{aligned}$$

That is, the ordinates whose heights are respectively 2·425 and 1·733 have as their abscissæ 3·37 and 2·08. By the property of the growth-curve the ordinate whose abscissa is 3·37 − 2·08 = 1·29 will have the height 2·425/1·733. Reference to the table shows that 1·29 is the logarithm of a number which lies between 1·4 and 1·5 and is approximately:—

$$1{\cdot}4 + 0{\cdot}1 \times \frac{1{\cdot}29 - 1{\cdot}28}{1{\cdot}55 - 1{\cdot}28} = 1{\cdot}404$$

We conclude that 2·425/1·733 = 1·40.

It will be found by arithmetic that the answers obtained are (to two decimal places) exactly correct in the first example and less than ¾ per cent in excess in the second example. Thus even the very simple table of p. 329 suffices to yield results of considerable accuracy.

§ 3. *Antilogarithms.*—As we have seen, a table of logarithms may be regarded as a list of the abscissæ corresponding to given ordinates of a given growth-curve. It would be equally easy to tabulate the ordinates corresponding to given abscissæ. Such a table is called "a table of antilogarithms"—the "antilogarithms" being the ordinates or "numbers" of the table of logarithms. Such a table is convenient when, at the end of a calculation conducted by means of logarithms, it is necessary to know the number corresponding to the logarithm finally obtained. The table on the next page consists of antilogarithms obtained from curve A of *Exercises*, fig. 50, and corresponds to the first column of the table on p. 329.

In § 2, example 1, the final logarithm was 0·756. The corresponding number or antilogarithm lies, then, between 1·17 and 1·23. By the method of proportional parts it is

$$1{\cdot}17 + 0{\cdot}06 \times \frac{0{\cdot}756 - 0{\cdot}6}{0{\cdot}8 - 0{\cdot}6} = 1{\cdot}217$$

ANTILOGARITHMS TO BASE 1·3.

log.	antilog (n).	log.	antilog (n).
0·0	1·00	1·0	1·30
0·2	1·05	1·2	1·37
0·4	1·11	1·4	1·44
0·6	1·17	1·6	1·52
0·8	1·23		

In example 2 the final logarithm was 1·29. The corresponding antilogarithm is

$$1{\cdot}37 + 0{\cdot}07 \times \frac{0{\cdot}09}{0{\cdot}20} = 1{\cdot}405$$

These results agree with the former ones to two decimal places.

[Ex. LIV, B, should now be worked.]

CHAPTER XXXIII.

COMMON LOGARITHMS.

A. Gunter's Scale and Logarithms obtained by Calculation (ch. XXIX., § 4 ; Ex. LV).

§ 1. *Logarithms apart from the Growth-curve.*—In the last three chapters Gunter's scale, the slide rule and tables of logarithms have all been considered in connexion with growth-curves. It is important to have a clear idea of the nature of this connexion. Suppose that a Gunter's scale were placed in your hands, but that you were quite ignorant of growth-curves and their properties ; what could you learn from an examination of the scale by itself? The answer is that the scale is a line so divided that the graduations at equal distances are always in the same ratio no matter what the distances may be nor where they are taken. For example, on the line GG of *Exercises*, fig. 50, any pair of graduations an inch apart have a ratio of 1·56, any pair 0·8 cm. apart a ratio of 1·15, and so on. (It is to be understood that the ratio meant is that of the right-hand number of the pair to the left-hand number.)

Similarly, if a table of antilogarithms were placed in your hands without any explanation of its mode of construction you might easily discover that there is always a constant ratio between pairs of numbers whose logarithms have a constant difference. For example, in the table on p. 332, if pairs of logarithms be selected whose difference is 1·0 (such as 0·0 and 1·0, 0·4 and 1·4, 1·8 and 2·8) the ratio of the second number (or antilogarithm) to the former is always 1·3. If the logarithms have a difference of 0·6 (e.g. 0·2 and 0·8, 1·4 and 2·0, 1·8 and 2·4) the ratio is always 1·17. If the difference between the logarithms is 1·2 the ratio of the numbers is 1·37, and so on.

A person who had discovered these facts could proceed to

make a Gunter's scale or a table of antilogarithms without the assistance of a growth-curve. He could, for example, make up his mind that the graduations which are separated by one-tenth of an inch on the scale, or the numbers whose logarithms differ by 0·1, should have a ratio of 1·01. By repeated multiplication by 1·01, unity being the original multiplicand, he could determine the graduations to be placed on the scale at points 0·1 inch, 0·2 inch, 0·3 inch, 0·4 inch, etc., from the beginning, or the numbers to be placed in the table against the logarithms 0·1, 0·2, 0·3, 0·4, etc. In order to complete the Gunter scale it would be necessary to pick out, by interpolation, the decimal graduations 1·1, 1·2, 1·3, etc., and to number them at convenient intervals as in *Exercises*, fig. 50. Similarly, to form a table of antilogarithms it would be necessary to calculate by the method of proportional parts, the antilogarithms corresponding to a series of equi-different logarithms and arrange them as on p. 332. By another application of the same method, a table of the logarithms corresponding to equidistant numbers could be calculated from the same figures.

It will now be seen that a growth-curve is by no means essential to the construction of a Gunter scale or of tables of logarithms and antilogarithms. The curve is useful merely because it offers an easy method of determining scale graduations and numbers which could otherwise be determined only by tedious and troublesome arithmetic. It will be understood, on the other hand, that the arithmetical method, though terribly laborious, can be carried out to any desired degree of accuracy, while the use of the growth-curve gives, like all graphical methods, results of strictly limited accuracy.

It should be noted that the factor by the constant repetition of which the Gunter scale or the table of logarithms is constructed need not be greater than unity. As a matter of fact in Napier's first table of logarithms it was less than unity. In such a table, as the numbers increase the logarithms must decrease. In the corresponding Gunter's scale the graduations would decrease towards the right. The corresponding growth-curve would slope downwards in the positive direction.

§ 2. *The Base of the Logarithms.*—In § 1 nothing has been said about the base of the logarithms. The reason is obvious. When the logarithms are obtained graphically from a growth-curve the base is simply the growth-factor used in drawing

the curve. It is selected, therefore, before the operation begins. When the curve is completed the base is the ratio between any two ordinates whose abscissæ differ by 1. In the table, then, the base is the ratio between any two numbers whose logarithms differ by 1. But if we begin by selecting arbitrarily some number which shall be the ratio between numbers whose logarithms differ by (say) 0·1 the base is not at first known. It must be found by taking the 10th power of the selected ratio. If we chose to work with a logarithm-difference of 0·01 the base would be the 100th power of the selected factor. In general, if we decide that numbers whose ratio is r shall have logarithms differing by $1/p$ of unity, then the base is r^p.

§ 3. *Derivation of Logarithms from a Gunter Scale.*—If we have a Gunter scale constructed by either the graphical or the arithmetical method it is easy to derive from it a table of logarithms to any given base. For example let the chosen base be 2. Lay the scale across squared paper as in fig. 79 so that the vertical graduated "1" meets the scale at the graduation "2". Then, since pairs of equidistant scale-graduations all have the same ratio, the ratio of the graduations lying on any pair of verticals 1 inch apart will be 2. If, then, the graduations of the verticals be regarded as logarithms and the scale-graduations as numbers we have the materials for a table in which, by the foregoing definition, the base will be 2. To compile the table we need only read off the graduations of the verticals which strike the edge of the scale at the points marked 1·1, 1·2, 1·3, etc. Similarly, if a table of logarithms is desired with 1·5 as base the Gunter scale must be swung round until the vertical graduated "1" meets its edge at the point marked "1·5". The logarithms can then be read off as before. Any other table can be constructed in the same way. To find the logarithms of numbers above 10 or below 1 the Gunter scale must, of course, be produced and the supplementary sections graduated in accordance with the principle studied in ch. xxxi., § 3.

[Ex. LV should now be worked.]

B. Common Logarithms (ch. xxix., § 6; Ex. LVI).

§ 1. *The Advantage of Logarithms to Base* 10.—Reflexion upon the principle used in A, § 3, suggests that the business of constructing a table of logarithms can be greatly facilitated by

adopting 10 as the base. Turn the Gunter scale round until the graduation "10" is on the vertical labelled "1". Then, as before, the logarithms of the numbers 1·1, 1·2 . . . 2·0, 2·1 . . . 9·8, 9·9, can be read off from the squared paper graduations by interpolation. They will all be decimal fractions between 0 and 1. But in this case it is unnecessary to produce the Gunter scale in order to read off the logarithms of numbers above 10 and below 1. For we know that if the scale were produced the graduation 100 would fall on the vertical labelled "2," and that the graduations 11, 12 . . . 20, 21 . . . 98, 99, would divide the scale between the verticals "1" and "2" in exactly the same way as the graduations 1·1, 1·2 . . . 2·0, 2·1 . . . 9·8, 9·9 divide the scale between the verticals "0" and "1". It follows that if the scale graduation n lies on the vertical whose graduation is l the scale graduation $10n$ lies on the vertical whose graduation is $+ 1 + l$. In other words, if the logarithm of a number n is l the logarithm of 10 times that number is $+ 1 + l$. By the same principle the logarithm of $100n$ will be $+ 2 + l$; of $1000n$, $+ 3 + l$; of $\frac{1}{10}n$, $- 1 + l$; of $\frac{1}{100}n$, $- 2 + l$; and so on. In general if p be any integer, positive or negative, the logarithm of $n \times 10^p$ is $p + l$.

§ 2. *Practical Use of Logarithms to Base* 10.—It follows from the foregoing that if 10 be chosen as base the only logarithms that need be determined are those of numbers between 1 and 10. The logarithm of a number N, greater than 10 or smaller than 1, can be derived from one of them by the following simple rule: (*a*) Express the number in the standard form (ch. XVII., § 8), $N = n \times 10^p$. (*b*) Find from the table the logarithm of n. This will be a decimal fraction l, less than 1. (*c*) Then $\log_{10} N = p + l$. Thus to find the logarithm of 743·6 to base 10 we begin by throwing the number into the form $7{\cdot}436 \times 10^2$. We then refer to a table and find that the logarithm of 7·436 is (to four places) 0·8713. We conclude that $\log_{10} 743{\cdot}6 = 2 + 0{\cdot}8713$ or 2·8713. If we required the logarithm of 0·07436 we should proceed similarly: we have $0{\cdot}07436 = 7{\cdot}436 \times 10^{-2}$; hence $\log_{10} 0{\cdot}07436$ is $- 2 + 0{\cdot}8713$. For compactness the logarithm is generally written $\bar{2}{\cdot}8713$, this form being adopted to indicate that the *minus* belongs to the whole number only, and that the decimal part of the logarithm is positive.

[Ex. LVI should now be worked.]

C. The Use of Tables (ch. xxix., § 6; Ex. LVII).

§ 1. *Tables of Logarithms.*—For all usual calculations logarithms to base 10 are so much the most convenient that they are universally used and are called "common" logarithms. They are printed in tables in which the logarithms of numbers between 1 and 10 are set out to various numbers of decimal places to meet different requirements of accuracy. The tables most used by engineers give the logarithms either to four or to five places. Navigators and astronomers usually require seven-figure logarithms. Seven-figure logarithms are necessary and sufficient also for most problems in finance and insurance, though for some such problems logarithms to twelve places are required. Tables carrying the logarithms to twenty and even sixty-one places have also been published, though they are rarely needed.

In using a table of common logarithms it is important to remember two things: (*a*) All the numbers given are to be taken as lying between 1 and 10, even though they are not generally so printed. Thus the numbers printed as 26, 782, etc., are to read as if they were printed 2·6, 7·82, etc., the decimal point being omitted to save space. (*b*) The logarithms are all decimal fractions less than 1. That is, the symbols "0·" ("nought point") must be read in front of every logarithm even if they are not printed. (*c*) In the logarithm of a number which is not between 1 and 10 the nought before the decimal point is replaced by the index of the power of 10 used in expressing the given number in the standard form, the *minus* sign being written above the index when it is negative.

The decimal part of the logarithm of a number is called its *mantissa*, the integral part its *characteristic*.

§ 2. *The Arrangement of Tables.*—The arrangement of the logarithms in a four-figure table will be understood from the following specimen:—

	0	1	2	3	4	1	2	3	4	5	6	7	8	9
3·1	·4914	·4928	·4942	·4955	·4969	1	3	4	6	7	8	10	11	12
3·2	·5052	·5065	·5079	·5092	·5106	1	3	4	6	7	8	9	11	12
3·3	·5185	·5198	·5211	·5224	·5238	1	3	4	5	6	8	9	10	12

The number 0·4914 in the row beginning 3·1 and the column headed 0 is the logarithm of 3·10. The logarithm of 3·12 is in the same row but in the column headed 2, i.e. 0·4942. Similarly log 3·14 is 0·4969. In the actual table the first set of columns is continued by others headed 5, 6, . . . 9. Thus the first complete row gives the logarithms of 3·10, 3·11, 3·12, . . . 3·19. The second set of (narrow) columns is called the "difference columns". They contain numbers which are in every case to be added to the last figure of the logarithm just mentioned. Thus to find log 3·136 we must take log 3·13, i.e. 0·4955 and add to it the number (8) in the difference column headed 6 :—

$$\log 3{\cdot}136 = 0{\cdot}4963$$

Similarly, to find log 3·245 we take log 3·24 = 0·5105 out of the second row and add to the last figure the 7 out of the difference column headed 5 :—

$$\log 3{\cdot}245 = 0{\cdot}5112$$

Example.—Find the logarithms of 3307 and 0·0003122. For log 3307 we have

$$3307 = 3{\cdot}307 \times 10^3$$
$$\log 3{\cdot}30 = 0{\cdot}5185$$
$$\log 3{\cdot}307 = 0{\cdot}5194$$
$$\therefore \log 3307 = 3{\cdot}5194$$

For log 0·003122 we have

$$0{\cdot}0003122 = 3{\cdot}122 \times 10^{-4}$$
$$\log 3{\cdot}12 = 0{\cdot}4942$$
$$\log 3{\cdot}122 = 0{\cdot}4945$$
$$\therefore \log 0{\cdot}0003122 = \bar{4}{\cdot}4945$$

After a little practice this process can, of course, be abbreviated.

It must be remembered that in most printed tables the decimal points given in the specimen are not inserted. The student should, however, always supply them when he copies the numbers and logarithms in calculations.

Five-figure logarithms are generally arranged upon a similar plan—as in the following specimen in which the decimal points are omitted.

	0	1	2	3	4	1	2	3	4	5	6	7	8	9
31	49136	49276	49415	49554	49693	14	28	42	56	70	83	97	111	125

Example 1.—Find log 3·146 to five places. Here we have

$$\log 3{\cdot}14 = 0{\cdot}49693$$
$$\log 3{\cdot}146 = 0{\cdot}49776$$

83 being added to the last two figures.

Example 2.—Find log 3·1065. This time we have to determine the difference corresponding to the last two figures, 65. If the number had been 3·106 we should have added the number in the 6th difference column (83); if it had been 3·107 the number in the 7th difference column (97). We must, therefore, actually add $83 + \frac{1}{2}(97 - 83) = 90$. Thus

$$\log 3{\cdot}10 = 0{\cdot}49136$$
$$\log 3{\cdot}1065 = 0{\cdot}49226$$

§ 3. *Antilogarithms.*—These tables of logarithms to n places can also be used to find to n significant figures the number or antilogarithm corresponding to a given logarithm. Nevertheless tables of logarithms to four or five places are generally accompanied by separate tables of antilogarithms. In such a table the "argument" is the mantissa of the logarithms, and is therefore generally printed as a decimal less than 1. The numbers in the table are numbers between 1 and 10 and must be read as if they had a decimal point after the foremost digit. When the number corresponding to the mantissa of the logarithm has been determined it must be multiplied by a power of 10 whose index is the characteristic of the logarithm. Thus if the given logarithm is 3·427 the table must be entered with the mantissa ·427. The antilogarithm or number of which this is the logarithm is (to three places) 2·673. The number whose logarithm is 3·427 is, therefore, $2{\cdot}673 \times 10^3$ or 2673. Similarly the antilogarithm of $\bar{3}{\cdot}427$ is $2{\cdot}673 \times 10^{-3}$ or 0·002673.

As far as concerns the difference columns, etc., the table of antilogarithms is arranged like the corresponding table of logarithms and is used in the same way.

Example.—Find from a table of five-figure antilogarithms the number whose logarithm is $\bar{2}{\cdot}31268$.

The antilogarithm of ·312 is found in the row beginning ·31 and in the column headed **2**. It is 2·0512. The next two figures lie between 60 and 70. The numbers in the difference-column headed **6** and **7** are 28 and 33 and their difference is 5.

The difference for 68 (= 60 + $\frac{4}{5}$ of 10) is therefore 28 + $\frac{4}{5}$ of 5, or 32. Thus we have:—

$$\begin{aligned} \text{antilog } 0{\cdot}31268 &= 2{\cdot}0512 + 0{\cdot}0032 \\ &= 2{\cdot}0544 \\ \therefore \text{antilog } \bar{2}{\cdot}31268 &= 2{\cdot}0544 \times 10^{-2} \\ &= 0{\cdot}020544 \end{aligned}$$

[Ex. LVII should now be worked.]

CHAPTER XXXIV.

THE LOGARITHMIC AND EXPONENTIAL FUNCTIONS.

(Ch. XXIX., § 7; Ex. LVIII.)

§ 1. *Summary of Previous Work.*—It will now be well to restate the theory of logarithms in a form which contains no reference to growth-curves. Choose any positive whole number p, greater than unity, and let $k = 1/p$. Choose also any positive number a, integral or fractional, and let $h = \sqrt[p]{a}$. Now construct (1) the complete arithmetic sequence whose starting term is zero and common difference k; (2) the complete geometric sequence whose starting term is 1 and common ratio h. Arrange these sequences so that the starting and other corresponding terms are against one another:—

$$\ldots, -3k, -2k, -k, \quad 0, \quad +k, +2k, \ldots \underset{[1]}{\mathbf{pk}}, \ldots$$

$$\ldots, \quad h^{-3}, \quad h^{-2}, h^{-1}, 1, \quad h^{1}, \quad h^{2}, \ldots \underset{[a]}{\mathbf{h^{p}}}, \ldots$$

Then each term of the arithmetic sequence is defined as the *logarithm* of the corresponding term in the geometric sequence, and a, the term in the geometric sequence which corresponds to 1 in the arithmetic sequence, is defined as the *base* of the logarithms.

Let mk and nk be any two terms of the arithmetic sequence —so that m and n are integers, positive or negative. Then, by the definition, $mk = \log_a h^m$, $nk = \log_a h^n$. By the same definition $(m \pm n)k = \log_a h^{m \pm n}$. Then since

$(m \pm n)k = mk \pm nk$, $h^{m+n} = h^m \times h^n$ and $h^{m-n} = h^m/h^n$

it follows that the sum of the logarithms of the two numbers h^m and h^n is equal to the logarithm of their product, while the difference of their logarithms is the logarithm of their quotient. Similarly, if r be any integer, $rm.k = \log_a h^{rm}$, that is $mk \times r = \log_a (h^m)^r$. Hence the logarithm of the rth

power of the number h^m is r times the logarithm of the number. If m/r is also an integer, mk/r and $h^{m/r}$ are corresponding terms of the two series, so that $mk/r = \log_a h^{m/r}$; that is the logarithm of the rth root of the number h^m is $1/r$th of the logarithm of the number. In this way the fundamental properties of the logarithm are established.

§ 2. *The Logarithmic Function.*—Now by making p exceedingly large it is possible to make k so small and h so little different from unity that the intervals between the terms in any part of either sequence become smaller than any specified number. That is to say, either of the sequences can be made to include, to any given closeness of approximation, any number that anyone chooses to mention, while the term 1 in the arithmetic sequence still has a as its corresponding term in the geometric sequence. Thus if a given number x is practically identical with a term of the geometric sequence, we can always find a term y corresponding to it in the arithmetic sequence. In other words, given any value of x there is always a y such that $y = \log_a x$.

In this way we reach the idea of a new function of x—the *logarithmic function*—whose properties are expressed by the relations—

$$\log_a x_1 x_2 = \log_a x_1 + \log_a x_2$$
$$\log_a (x_1/x_2) = \log_a x_1 - \log_a x_2$$
$$\log_a x^n = n \log_a x$$
$$\log_a \sqrt[n]{x} = \frac{1}{n} \log_a x$$

All these properties may be briefly summed up in the following statements. (i) If a variable y depends on a variable x in such a way that to two values of x with a fixed ratio there always correspond two values of y with a fixed difference, and if $y = 0$ when $x = 1$, then y is said to be the logarithm of x. (ii) If, when the ratio between two values of x is a, the difference between the corresponding values of y is 1, a is said to be the base of the logarithms.

§ 3. *The Antilogarithmic Function.*—In the foregoing argument the term of the geometric sequence has been taken as the independent variable. Instead of doing so we may regard the term of the arithmetic sequence as the independent variable —that is, we may put $x = mk$ and $y = h^m$. In that case we can, of course, write $y = \text{antilog}_a x$. The function expressed

by the symbolism $\text{antilog}_a x$ is obviously the inverse of the logarithmic function with the same base. It may be called the *antilogarithmic function.*

The properties of the antilogarithmic function can be found by the method already applied to the logarithmic function. Let mk and nk be any two values of x (just as in § 2 they were values of y). Then the corresponding values of y are h^m and h^n; that is

$$h^m = \text{antilog}_a(mk)$$
$$h^n = \text{antilog}_a(nk)$$

Now we have

$$\begin{aligned}\text{antilog}_a(mk) \times \text{antilog}_a(nk) &= h^m \times h^n \\ &= h^{m+n} \\ &= \text{antilog}_a(mk + nk)\end{aligned}$$

That is to say, if any two values of x are taken, x_1 and x_2,

$$\text{antilog}_a x_1 \times \text{antilog}_a x_2 = \text{antilog}_a(x_1 + x_2)$$

In the same way it can be shown that

$$\text{antilog}_a x_1 \div \text{antilog}_a x_2 = \text{antilog}_a(x_1 - x_2)$$

§ 4. *Fractional Indices.*—The most notable thing about the properties of the antilogarithmic function is that the values of x—no matter whether they are integral or fractional, positive or negative—are combined in multiplication and division exactly as if they were indices or "exponents" of powers of the same number. For this reason the function is often called the "exponential function," the symbolism $y = \exp_a x$ being used instead of $y = \text{antilog}_a x$.

Now such symbolism as $a^{x_1} \times a^{x_2} = a^{x_1 + x_2}$ is both much more easily written and has become, by use, much more familiar than either

$$\text{antilog}_a x_1 \times \text{antilog}_a x_2 = \text{antilog}_a(x_1 + x_2)$$

or

$$\exp_a x_1 \times \exp_a x_2 = \exp_a(x_1 + x_2)$$

and is also more easily expressed in words. It would for these reasons alone be worth while to inquire whether the antilogarithmic (or exponential) function of x in which values of x are certainly *combined* as if they were indices cannot be *written* as if they were indices. The following argument shows that they may be so written.

If from the arithmetic sequence on p. 341 each pth term be picked out, starting from 0 and counting both ways, we obtain the new arithmetic sequence, . . . − 4, − 3, − 2, − 1, 0, + 1, + 2, + 3, + 4, . . . while the corresponding terms of

the original geometric sequence form the new geometric sequence . . . a^{-4}, a^{-3}, a^{-2}, a^{-1}, 1, a^1, a^2, a^3, a^4 . . .

Remembering that y represents terms of the geometric and x terms of the arithmetic sequence, we see at once that *for these terms* the function $y = \text{antilog}_a x$ or $y = \exp_a x$ can be expressed in the alternative form $y = a^x$. It is, therefore, a natural suggestion that the symbolism $y = a^x$ should be used as an alternative way of expressing the relation between *all* values of x and the corresponding values of y when y is the antilogarithm of x.

If we adopt this plan we must, of course, abandon the old definition of an index. The expressions $a^{3\cdot7}$ and $a^{-8\cdot2}$ cannot mean that unity is repeatedly multiplied by a 3·7 times in the first case and divided by a 8·2 times in the second, for these operations are impossible. But such a consideration need not prevent us from adopting the new symbolism if we wish to do so. We can *re-define* a^x as meaning "the antilogarithm of x to the base a". When x is a whole number it will have, in addition to this meaning, the old meaning of repeated multiplication or division. When it is not a whole number it will have the new meaning only. There could, in fact, be only one fatal objection to the proposed practice. We could not adopt it if antilogarithms did not combine with one another in accordance with the laws of indices. Since they do so combine there is no reason why the symbolism should not be used whenever it proves convenient.

The effect of these arguments is best understood by means of examples. On p. 329 we read that $\log_{1\cdot3} 1\cdot1 = 0\cdot36$ and $\log_{1\cdot3} 1\cdot4 = 1\cdot28$. From these results, or from the table on p. 332, we derive statements which can be expressed in three equivalent ways :—

$$\begin{aligned} \text{antilog}_{1\cdot3}\, 0\cdot36 &= 1\cdot1 & \text{antilog}_{1\cdot3}\, 1\cdot28 &= 1\cdot4 \\ \exp_{1\cdot3}\, 0\cdot36 &= 1\cdot1 & \exp_{1\cdot3}\, 1\cdot28 &= 1\cdot4 \\ (1\cdot3)^{0\cdot36} &= 1\cdot1 & (1\cdot3)^{1\cdot28} &= 1\cdot4 \end{aligned}$$

Employing the last form as the most convenient, we have

$$\begin{aligned} 1\cdot4 \times 1\cdot1 &= (1\cdot3)^{1\cdot28} \times (1\cdot3)^{0\cdot36} \\ &= (1\cdot3)^{1\cdot64} \\ &= 1\cdot54 \text{ by either of the tables.} \end{aligned}$$

Similarly

$$\begin{aligned} 1\cdot4 \div 1\cdot1 &= (1\cdot3)^{1\cdot28} \div (1\cdot3)^{0\cdot36} \\ &= (1\cdot3)^{0\cdot92} \\ &= 1\cdot27 \text{ by either of the tables.} \end{aligned}$$

Again $(1{\cdot}1)^3 = \{(1{\cdot}3)^{0{\cdot}36}\}^3$
$= (1{\cdot}3)^{1{\cdot}18}$
$= 1{\cdot}36$

Finally $\sqrt[4]{(1{\cdot}4)} = \sqrt[4]{\{(1{\cdot}3)^{1{\cdot}28}\}}$
$= (1{\cdot}3)^{0{\cdot}32}$
$= 1{\cdot}09$

All these (approximate) results can be confirmed by arithmetic.

§ 5. *Graphs of the Functions.*—It will now be useful to return to the growth-curve. As we saw long ago the ordinates of this curve may be regarded as "numbers" or "antilogarithms" and the corresponding abscissæ as "logarithms". Given a table of logarithms or of antilogarithms the curve might be drawn without any reference to the growth-problems which actually suggested it. In other words, it is the graph of the antilogarithmic or exponential function. It may, therefore, conveniently be called the *antilogarithmic* or *exponential curve.* The second of these names is the one generally employed.

When the growth-curve was first drawn it could not be said to be the graph corresponding to any known formula. Some of its points were given by the formula $y = a^x$, but only those where x is integral. But by § 4 we can now say that *every* point on the curve is given by the formula $y = \text{antilog}_a x$, or $y = \exp_a x$ or (lastly) $y = a^x$. The last, as the most concise and familiar, will henceforward be taken as the standard formula of the curve.

To obtain the graph of the logarithmic function it is sufficient (as in the case of all functions of which one is the inverse of the other) to turn the exponential curve (1) through 180° about the x-axis, (2) through 90° anticlockwise about an axis through the origin perpendicular to its plane.

[Ex. LVIII should now be worked.]

CHAPTER XXXV.

NOMINAL AND EFFECTIVE GROWTH-FACTORS.

(Ch. XXIX., § 8; Ex. LIX.)

§ 1. *"Nominal" and "Effective" Rates of Growth.*—In this chapter we are to study in detail one of the most important differences between the modes of uniform growth contrasted in ch. LI.

Suppose it to be known that a quantity, of present magnitude Q_0, is growing in accordance with the first law. Suppose, also, that its changing magnitude has been observed during $1/n$th of the time-unit and that during this interval each unit of magnitude changes by an amount a'. Then the growth-difference for a complete time-unit will, of course, be $a' \times n$. Putting $a = a' \times n$ the formula for the magnitude of the quantity at time t becomes

$$Q = Q_0 (1 + at)$$

Conversely, if we are given that the growth-difference of unit magnitude during unit time is a, then we know that the growth of a unit during *any* $1/n$th of this term will be a/n and the change in the magnitude of the whole quantity $Q_0 a/n$.

In the case of the second form of uniform continuous growth there is no such simple connexion between the changes of magnitude during the whole time-unit and one of its fractions. The difference will be best brought out by an example. Make the assumption that the population of a certain town increases with a constant growth-factor, and that on 1 January this year it was 40,000. Suppose, also, that at the end of the first quarter the population was recounted and found to be 40,260. Then there has been during this quarter an actual increase of 260 or $260/40 = 6{\cdot}5$ per thousand of the population on 1 January. If the same actual increase occurred in each quarter the growth-difference

for the year would be $6{\cdot}5 \times 4 = 26$ persons per thousand. We may say, then, that during the first quarter the observed increase was *at the rate of* 26 per thousand per annum.

But this number does not measure the actual yearly increase in the population. An increase from 40,000 to 40,260 in one quarter means a quarterly growth-factor of $1{\cdot}0065$ and therefore an annual growth-factor[1] of $(1{\cdot}0065)^4 = 1{\cdot}02625$. Thus the actual annual increase would be 26·25 per thousand.

These facts may be shortly expressed by the statement that while the *nominal* rate of increase during this quarter was 26 per thousand per annum, the *effective* rate of increase was 26·25 per thousand per annum.

It is obvious that in an actual town accidental causes—immigration, emigration, etc.—are likely to affect the population seriously in so long a time as a quarter. If we wanted to know the natural rate of change of the original 40,000 it would be necessary to consider the increase during a shorter period. Suppose, then, that during the first fortnight of the year the population increases from 40,000 to 40,040. This is an increase of 1 per thousand for the fortnight or an increase at the *nominal* rate of 26 per thousand per annum. But though the nominal rate of increase is the same as before the *effective* rate is different. For a fortnightly growth-factor $1{\cdot}001$ implies an annual growth-factor $(1{\cdot}001)^{26} = 1{\cdot}02633$. That is, the effective rate of increase is 26·33 per thousand per annum.

In general, if i is the observed increase per *unit* magnitude during the first $1/n$th of the unit of time, then the growth-factor for the complete unit of time is $(1 + i)^n$. The nominal increase of unit magnitude for the unit of time is $j = ni$. The effective increase is

$$(1 + i)^n - 1 = (1 + j/n)^n - 1$$

§ 2. *Financial Applications.*—These ideas have important financial applications. Suppose that a sum of £1 is deposited in a bank which gives interest at the rate of j per *pound* per annum. Suppose, also, that the interest is paid once a year and is on each occasion of payment added to the deposited principal. Then the original £1 will increase by annual "jerks" with a growth-factor of $(1 + j)$. Now suppose that the interest is paid quarterly. That is, suppose that at the

[1] Calculated by means of seven-figure logarithms.

end of the first quarter the banker adds $\frac{1}{4}j$ to your £1 and gives interest during the second quarter on the whole sum $1 + j/4$. Then in this case your principal will increase by quarterly "jerks" showing a growth-factor of $1 + j/4$ for the quarter and $(1 + j/4)^4$ for the year. Thus while the nominal rate of interest is j per pound per annum the effective rate is $(1 + j/4)^4 - 1$ per pound per annum. In general, if interest at the nominal rate of j per pound per annum is added to the principal n times a year the effective rate of interest is $i = (1 + j/n)^n - 1$.

§ 3. *Graphic Determination of Rates.*—It is easy to exhibit graphically the effective rate which corresponds to a given nominal rate of growth. Let AP (fig. 83) be any exponential curve, and AO the ordinate of unit magnitude. Let j be the nominal rate of increase per unit time, the actual increase being j/n during the first $1/n$th of a time-unit. Find the

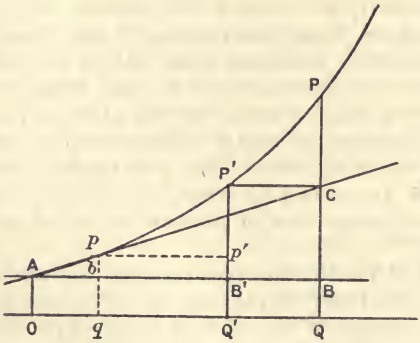


FIG. 83.

ordinate P′Q′ whose height is $1 + j$, and draw AB′B horizontally. In P′B′ take $B'p' = 1/n$ of P′B′ and draw $p'p$ horizontally to meet the curve in p. Then

$$pq = p'Q' = (1 + j/n)$$

Draw the secant ApC and let it cross the horizontal through P′ in C. Through C draw the ordinate PQ. Then in the similar triangles, Apb, ACB, $CB = P'B' = p'B' \times n = pb \times n$. Hence $AB = Ab \times n$, and (by the fundamental property of the exponential curve)

$$PQ = (pq)^n = (1 + j/n)^n$$

Thus PB = $(1 + j/n)^n - 1$ is the effective rate of increase which corresponds to a nominal rate of increase P′B′.

§ 4. *The Case of Continuous Increase.*—The foregoing argument applies equally well to a continuously growing quantity (e.g. rising temperature) whose increase is *measured* at the end of the first nth of the time-unit, and a quantity (such as the principal deposited in a bank) which actually *acquires* its increments only at intervals of $1/n$th of the time-unit. This fact enables us to understand how the distinction between the nominal and the effective rate of interest can still be used when growth by "jerks" passes into continuous growth. Many important financial calculations depend upon the supposition that interest at the rate of j per pound per annum is added to the principal as fast as it is earned. In such cases the growth of the principal is represented graphically not by a series of unequal steps (of which Abp, fig. 83, is the first) but by the exponential curve itself, every ordinate representing the amount of the principal at some moment. The transition from the case in which the interest is "converted into principal" at n distinct and equidistant times in the year to the case in which the conversion takes place every moment may be supposed to be effected by making n become larger endlessly. Accompanying the increase in n the point p' (fig. 83) will approach B′ and the point p move along the curve towards A. At the same time the secant ApC will approach and eventually become indistinguishable from the tangent AC (fig. 84), but it will never pass beyond it. Thus when n is so large that the interest may be considered as converted into principal every moment, the effective rate of interest, PB (fig. 84), is obtained from the nominal rate P′B′ by a construction which differs from that of fig. 83 only by the substitution of the tangent at A for the secant ApC. We may express the result symbolically by the statement that, if P′Q′ (fig. 84) = $1 + j$, then PQ = $(1 + j/n)^n$ when $n = \infty$.

§ 5. *The Meaning and Use of "e".*—When n is a definite number the value of the effective rate which corresponds to a given nominal rate can, of course, be calculated. It is simply a question of computing $(1 + j/n)^n$ by means of a table of logarithms. But, when all we can say about n is that it is endlessly great, computation in accordance with this formula becomes impossible. Our last task is to seek some way of calculating the effective rate in such cases.

In fig. 84 let $P'B' = j$ as before, and let $R'D' = 1$; that is, let $R'S' = 2$. Then by § 4, if $n = \infty$, the ordinate $RS = (1 + 1/n)^n$ and the ordinate $PQ = (1 + j/n)^n$. Now in the similar triangles AED, ACB, $ED = R'D' = 1$ and $CB = P'B' = j$. Hence $AB = AD \times j$ and $PQ = (RS)^j$. That is to say

$$(1 + j/n)^n = \{(1 + 1/n)^n\}^j \text{ when } n = \infty \quad . \quad \text{(i)}$$

It is most important to note here that the height of RS must always be the same no matter what exponential curve is used to find it. To assure yourself of this return to fig. 83 and suppose that $P'B' = 1$, and that $p'B'$ is $1/n$, n being, of course, a definite number. Then since $PQ = (1 + 1/n)^n$

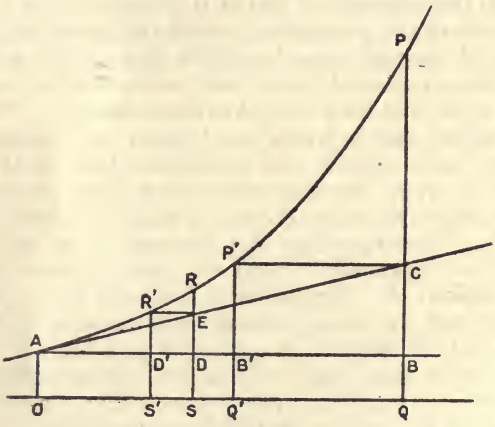


FIG. 84.

where n is a definite number, its value is quite independent of the growth-factor of the curve. Since this is true however large n is, so long as it is a definite number, it must be true when n has become so large that AC is henceforward indistinguishable from the tangent. That is, it is true when n is endlessly large or $n = \infty$. It follows that although RS (fig. 84) would appear in different positions in different exponential curves its height has the same value in all curves. In other words, when $n = \infty$ $(1 + 1/n)^n$ has a definite numerical value.

It is customary for conciseness to symbolize this value by

the letter e. The result (i) above can then be expressed in the simple form

$$(1 + j/n)^n = e^j \quad . \quad . \quad . \quad \text{(ii)}$$

while the effective rate of increase in the case when increase takes place every moment at a nominal rate of j per unit time can be written $i = e^j - 1$. The value of e will be found, by careful drawing and measurement, to be approximately 2·72. Algebraic methods, to be described later, show that it is 2·71828 . . . Like π it is an endless decimal which has been calculated to several hundreds of places.

[Ex. LIX may now be worked].

EXERCISES SUPPLEMENTARY

TO SECTIONS II AND III.

SUPPLEMENTARY EXERCISES.

*** The numbers in ordinary type refer to the pages of *Exercises in Algebra*, Part I; the numbers in heavy type to the pages of this book.

EXERCISE		PAGES
LX.	The Use of Logarithms in Trigonometry	303, 355
LXI.	Polar Co-ordinates	308, 356
LXII.	Some Important Trigonometrical Identities	310, 359
LXIII.	The Parabolic Function	317, 360
LXIV.	Implicit Quadratic Functions (I) . .	320, 361
LXV.	Implicit Quadratic Functions (II) . .	325, 362
LXVI.	Mean Position	331, 366
LXVII.	Root-mean-square Deviation . . .	337, 369
LXVIII.	The Binomial Theorem	340, 372
LXIX.	The Generalization of Wallis's Law . .	349, 374

CHAPTER XXXVI.

THE PROGRAMME OF EXERCISES LX-LXV.

§ 1. *The Contents of the Section.*—The exercises of this section are supplementary in the sense that they introduce no new algebraic notation and no new fundamental idea; their contents are in all cases fairly straightforward developments of the subject-matter of previous exercises. It is chiefly for this reason that the preliminary discussions which have formed the bulk of previous sections of this book are now discontinued, and the student left to work his way through the exercises with the aid of a few "Notes". There is, however, another reason for this change in procedure. We have now to deal with a student of some maturity who should soon be called upon to acquire by unaided reading the information he needs in order to do his work. The practice adopted in this supplementary section aims at preparing him to exercise (in Part II) this fuller responsibility for his own progress.

The nature of the section makes any division into subsections rather arbitrary. There is, however, a certain continuity in Exs. LX-LXV, due to the fact that they all deal directly or indirectly with the properties and uses of the trigonometrical ratios. Similarly, Exs. LXVI-LXIX deal more or less with applications and further developments of Wallis's Law. There will, therefore, be a certain convenience in considering these two groups of exercises in separate chapters.

§ 2. *The Use of Logarithms in Trigonometry* (Ex. LX).—The first exercise exemplifies the use of logarithms in evaluating trigonometrical formulæ. As we have seen, the desire to make these evaluations a less burdensome business was the motive which led Napier to the invention of his "artificial numbers". One of the evil consequences of teaching trigonometry as a distinct subject has been that an excessive amount

of importance has been attached to these computations. The single short exercise here given to them probably represents the amount of consideration to which they are entitled in a general, as opposed to a technical, course of mathematics. It will be observed that no attempt is made to classify the different problems formally into "cases".

Division A is intended simply to give practice in using tables of logarithms of the functions. In connexion with these examples the teacher is recommended to follow the growing custom of using logarithms in which the characteristic has its proper value instead of being increased by 10. For the ordinary pupil and for ordinary purposes the intrusive 10 of the "tabular logarithm" is simply an occasion of stumbling and has nothing whatever to commend it. An incidental difficulty in the use of logarithms of the trigonometrical functions is that of dealing with negative values. **Nos. 5, 6** are intended specially to give practice in overcoming it, the proper procedure in these cases being explained in a note prefixed to No. 5.

The examples of division B cover familiar ground. The identities of **Nos. 10-14** are developments of an argument begun in Ex. LX and here to be repeated and brought to its conclusion. The note before **No. 16** suggests that the student shall be contented to commit to memory only the expression for calculating $\cos \alpha$ in terms of the sides of the triangle, this particular one of the three allied identities being chosen simply because it is the easiest to remember. The artificial questions which required the candidate to be ready with all three in order to effect a solution by means of logarithms selected by the examiner have fortunately disappeared from the modern examination paper.

Lastly, in division C, a few examples are given on the formula usually called by continental writers "Mollweide's equation". The geometrical argument suggested by the construction in *Exercises*, fig. 51, is much easier for the average student than the algebraic proof current in English textbooks. Familiarity with the use of Mollweide's identity should be regarded as a luxury rather than a necessity but will probably continue to be expected of candidates in examinations.

§ 3. *Polar Co-ordinates* (Ex. LXI).—The device of fixing the position of a point by means of an angular measurement and a distance has often been illustrated in the exercises from

Ex. II onwards. The present short exercise is intended to show the power of the method as a means of geometrical description, and to prepare the student to use it freely in the important arguments of Ex. LXV and in many later exercises.

The path of Halley's comet, **No. 2**, is really a very flat ellipse, but the given formula, as is shown in Ex. LXIII, No. 13, describes the portion near the sun as a parabola. In drawing the graphs in **Nos. 3-9** different members of the class should assign different values to the constants. Fig. 85 shows the limacon, **No. 4**, when $a = 2$ and $b = 1$. This figure degenerates into the cardioid, **No. 3**, when $b = 2$ also. Fig. 86 shows the "three-leaved rose" described by the polar

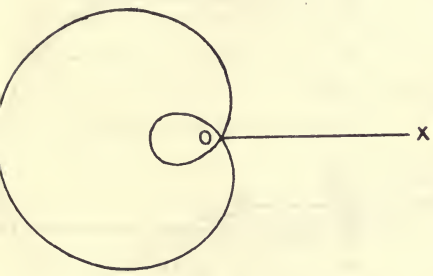


FIG. 85.

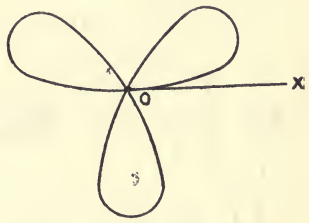


FIG. 86.

formula $r = a \sin 3a$, **No. 5**. The one corresponding to $r = a \cos 3a$ is of the same form rotated anticlockwise through 90°. In drawing these graphs, as well as those of No. 6, it must be remembered that the students have as yet learnt to measure angles only up to 360°. Thus the greatest value which can be assumed by a in No. 5 is 120°. It will be found, however, that this range of the variable is sufficient to give the whole of the curve. For example, in plotting $r = a \sin 3a$ as a passes through the value 60°, $3a$ passes through the value 180° and the radius vector begins to be negative. Thus between $a = 60°$ and $a = 120°$ the lower loop of the "rose" will be traced. The loop in the second quadrant will be obtained by giving to a *negative* values from 0° to − 60°. As a passes through the values from − 60° to − 120°, $3a$ will move from − 180° to − 360°. Thus the values of sin $3a$ and therefore of r will once more be positive

and the lower loop will be traced out a second time. There is, of course, no objection to using this graph as a means of introducing the idea of angles greater than 360°, but, on the whole, it is better to delay this important step forward until (as in Section VII) a more substantial justification for it appears.

Fig. 87 shows the conchoid, **No. 9**, with $a = 1, b = 2$. If $b = -2$ the figure will be reversed from right to left; if $b = a$ the loop will degenerate into a point.

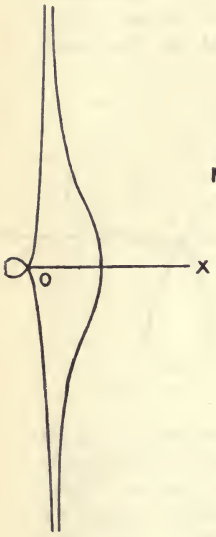


Fig. 87.

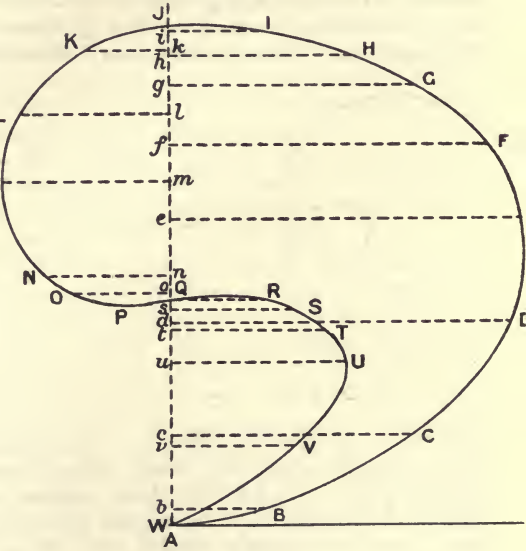


Fig. 88.

No. 10, which illustrates the extraordinary movements of which the Australian boomerang is capable, is likely to prove very interesting to boys, if not to their sisters. The data have been obtained by copying (roughly) a figure given by Mr. G. T. Walker in his elaborate paper on the subject in the *Philosophical Transactions*, vol. 190, A. Fig. 88 gives the "bird's-eye" view of the flight of the boomerang; fig. 89 shows how its movement would appear if watched by an observer at a considerable distance on a line bisecting AJ at right angles. The vertical scale is, of course, exaggerated.

To obtain fig. 89 perpendiculars B*b*, C*c*, D*d*, etc., are drawn to AJ in fig. 88 and the distances A*b*, *bc*, etc., repeated along the base line of fig. 89. At the points *a*, *b*, *c*, etc., lines A′*a*, B′*b*, C′*c*, etc., are drawn to represent to scale the values of *h* given in the table. A smooth curve is drawn through A′B′C′ . . . .

§ 4. *The Sum and Difference Formulæ* (Ex. LXII).—In the discussion which preceded Ex. L (ch. XXVIII., B, p. 295) it was necessary to investigate a formula for calculating sin $(\alpha + \beta)$ when the sines and cosines of α and β are known. The simple proof there given was adequate to the purpose, but applies only in the cases where $\alpha + \beta$ is less than 90°. Ex. LXII takes up the whole subject of the "sum and difference formulæ" for two angles, including all cases in which their

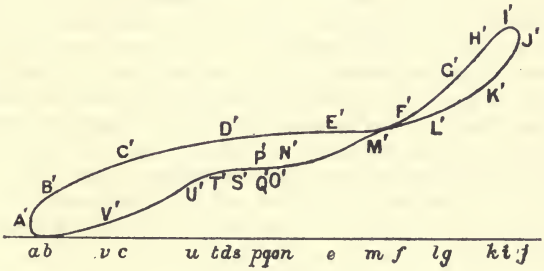


FIG. 89.

sum is not greater than 360°. The method followed starts from the notions of a "vector" and its resolution into "components" at right angles—notions with which the student made acquaintance as early as Ex. XIX. The treatment is straightforward and needs no elucidation. The teacher will observe that the mode of attack chosen makes it simpler to deal with the factorization-formulæ for cos α + cos β, etc., before the identities more specially known as the sum and difference formulæ. It will be noted, also, that only sufficient examples are given to familiarize the student with the results themselves and with the few transformations of constant practical importance which they are needed to effect. The formulæ will shortly be put to an important use in Ex. LXV.

§ 5. *Quadratic Functions.*—Exs. LXIII-LXV may be con-

veniently considered together since they all deal with functions of the second degree and may be regarded as continuing the discussions of Exs. XLII-XLVI.

Ex. LXIII is a short exercise dealing entirely with the parabolic function. Division A illustrates a type of elimination-process which has often to be used in dynamics and physics; division B explores in a simple way the connexion between the Cartesian and polar formulæ which describe a parabola.

In **No. 1** the substitution of $d/12$ for t gives the formulæ

$$h = 6d - d^2/9 \quad . \quad . \quad . \quad . \quad \text{A}$$
$$= 81 - \tfrac{1}{9}(27 - d)^2 \quad . \quad . \quad . \quad \text{B}$$
$$d = 27 \pm 3\sqrt{(81 - h)} \quad . \quad . \quad . \quad \text{C}$$

From B we see that the greatest height reached above the point of projection is 81 feet, or, above the ground, 86 feet. Substitution of 30 (i.e. 35 − 5) for h in C gives $d = 48\frac{3}{7}$ feet and $t = 48\frac{3}{7}/12 = 4\frac{1}{28}$ seconds. These are respectively the values of the horizontal range and the time of flight—the former of these two terms meaning the distance between the vertical through the hand of the thrower and the vertical through the point where the ball hits the roof. For the second part of the question we must substitute $-14 - 5 = -19$ for h in C, whence $d = 57$ feet and $t = 4\frac{3}{4}$ seconds.

In **No. 2** the formula for the line of greatest slope is evidently

$$h = \tfrac{1}{30}d - 5 \quad . \quad . \quad . \quad \text{D}$$

Substituting this value for h in A we obtain the quadratic equation

$$10d^2 - 537d - 450 = 0$$

which gives $d = 54{\cdot}52$. From D we now have $h = -3{\cdot}18$. These are the required co-ordinates of the point where the ball hits the ground. Let R be the range measured along the slope of the hill. Then we have

$$\text{R}^2 = d^2 + h^2$$
$$= d^2(1 + \tfrac{1}{900})$$

whence $\text{R} = d(1 + \frac{1}{1800})$

to a sufficient degree of accuracy. In the present case this argument gives the result $\text{R} = 54{\cdot}55$; that is, the slope of the hill adds less than half an inch to the range of the ball.

When the ball is thrown downhill (**No. 3**) we must replace D by the formula

$$h = -\tfrac{1}{30}d - 5 \quad . \quad . \quad . \quad \text{E}$$

which leads to the quadratic equation

$$10d^2 - 543d - 450 = 0$$

with 55·12 feet as the value of d. In this case R, determined as before, is 55·15 ft.

In all these calculations only the positive values of the roots have been retained. There is no difficulty in seeing that the negative roots do not apply to the problems proposed.

In **No. 13** the two formulæ are seen to be equivalent by the consideration that values of a differing by 180° would make their denominators identical. In order, therefore, that the same value of a may lead to the same point in each case the two fractions must be prefixed by opposite signs. If p is positive the parabola will be "head down"; if negative, "head up".

In all previous exercises functions of x have been "explicit"; that is, they have been expressed by formulæ in which the subject is y. Exs. LXIV, LXV introduce the study of "implicit" functions of x and y and of the corresponding graphs. The observations made in ch. IV., § 9, are here particularly relevant. Our business in these exercises is not to study "the equations of the conic sections," but to use our knowledge of these graphic forms to illuminate the properties of implicit quadratic functions of x and y.

Throughout Ex. LXIV constant use is made of the principle established in Ex. XXXIX, p. 218, namely, that if a graph is moved horizontally through a distance a and vertically through a distance b its formula must be amended by the substitution of $x - a$ for x and $y - b$ for y. Thus in **No. 2** the function of which the circle in its new position is the graphic expression is

$$(x + 3)^2 + (y - 4)^2 = 16.$$

Conversely, since the function in **No. 4** (iv) can be written

$$(x + 7)^2 + (y - 4)^2 = 64$$

its graph is a circle of radius 8 whose centre is at the point $(-7, +4)$. The same principle and its converse are, in **Nos. 22-4**, applied to the ellipse, and the corresponding functions.

Fig. 90 illustrates **No. 12**, the line PQ being the graph of the equation

$$4x - 6y - 13 = 0.$$

Since the points P and Q are on both circles their co-ordinates must satisfy both the equations

$x^2 + y^2 - 25 = 0$ and $x^2 + y^2 - 4x + 6y - 12 = 0$ and will therefore satisfy an equation in which the left-hand side is the difference between the left-hand sides of these two. That is, it will satisfy

$$4x - 6y - 13 = 0.$$

It follows that the line which is the graph of this equation is the "radical axis" or line through the intersections of the two circles.

Division B is very important in view of later applications, but requires little comment. All the examples should be worked, special attention being given to **Nos. 15-19.** The ellipse of **No. 19** is, of course, simply that of No. 16 rotated through a right angle about the origin.

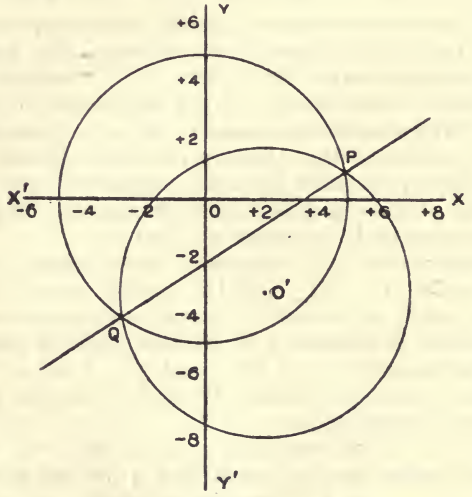


FIG. 90.

Previous exercises have given numerous illustrations of the fact that a single graph in different positions may correspond to many different functions. Hitherto all the different positions of the graph have been derivable from one another by vertical and horizontal movements and our repertory of functions studied has been limited by this condition. In Ex. LXV the restriction is removed, and we enter upon the study of functions which correspond to the now familiar graphic forms

after they have been not only displaced vertically and horizontally from their standard positions but in addition rotated about the origin. In the course of the argument it is shown that every possible function of the second degree corresponds to one of the standard graphic forms in one of the positions which these movements make possible. Thus the study of these transformations gives occasion for an exhaustive review of quadratic functions, explicit and implicit. It is well, however, to remind the student that the method here employed does not possess universal validity; we have already seen (p. 288) that all cubic functions do *not* correspond to the graph of $y = ax^3$ in the various positions which it may be made to assume.

The geometrical principle upon which these transformations depend is explained in the note at the beginning of the exercise and should be thoroughly understood. In **No. 1** we may suppose any (i.e. every) point of the graph to be joined with the origin. It is then obvious that if one of the joining lines is rotated through an angle α all will be, so that the relation between the old and the new co-ordinates established in the note will hold good for every point of the graph. **No. 2** gives an interesting and striking verification of this conclusion.

The first important application of the new principle of transformation occurs in **No. 5**. Here, substituting $(x - y)/\sqrt{2}$ for x and $(y + x)/\sqrt{2}$ for y in the standard rectangular hyperbolic function, we obtain $x^2 - y^2 = a^2$ as the formula for a rectangular hyperbola in the position lettered P′ in fig. 91. The next few examples are intended to bring out the striking and important analogies between this hyperbolic function and the circular function $x^2 + y^2 = a^2$ which are to reach their logical conclusion in the theory of the hyperbolic sine and cosine in Section VII. Fig. 91 is the diagram which the student is, in these examples, instructed to build up. The curves drawn with firm lines are the circle and the rectangular hyperbola which corresponds to it; the broken curves represent the ellipses which can be derived from the circle and the hyperbolas derived in the same way (**No. 12**) from the rectangular hyperbola.

The examples of division B give exercise in finding the functions which correspond to the parabola, ellipse and hyperbola after the specified displacements, by translation and

rotation, from their standard positions. In **No. 22** the original relation is

$$x^2 - y^2 = \tfrac{25}{4}$$

in accordance with the note before No. 13, and the cosine and sine of the angle whose tangent is $\frac{1}{3}$ are respectively $3/\sqrt{10}$ and $1/\sqrt{10}$. When the hyperbola is rotated anticlockwise

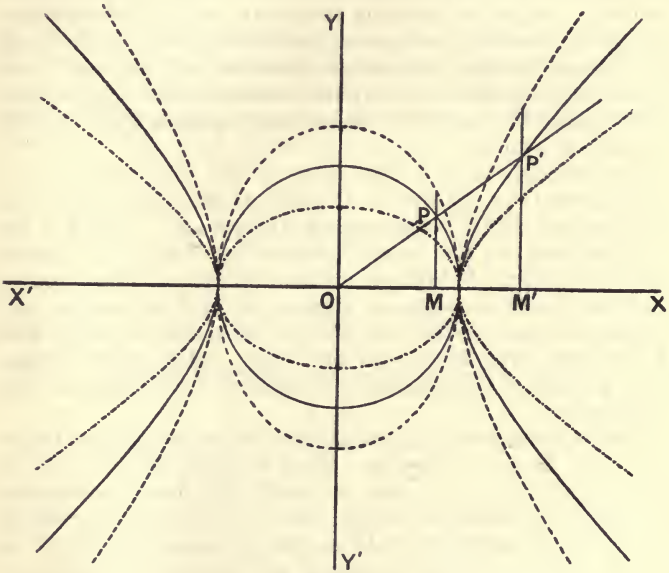


FIG. 91.

through this angle from its standard position it corresponds to the relation

$$\frac{(3x + y)^2}{10} - \frac{(3y - x)^2}{10} = \frac{25}{4}$$

that is

$$16x^2 + 24xy - 16y^2 = 125.$$

No. 26 is an example of the converse problem. If the graph is rotated *clockwise* through an angle α we have (as in No. 1) that it now corresponds to the implicit function

$$3(x\cos\alpha - y\sin\alpha)^2 + 8(x\cos\alpha - y\sin\alpha)(y\cos\alpha + x\sin\alpha) - 3(y\cos\alpha + x\sin\alpha)^2 = 10$$

which can be thrown into the form

$$(x^2 - y^2)(3\cos^2\alpha + 8\cos\alpha\sin\alpha - 3\sin^2\alpha) + xy\{8(\cos^2\alpha - \sin^2\alpha) - 12\cos\alpha\sin\alpha\} = 10.$$

But if the axis of the hyperbola is now coincident with the x-axis the term involving xy should disappear. The condition for this disappearance is

$$4(\cos^2\alpha - \sin^2\alpha) = 3 \times 2\cos\alpha\sin\alpha$$

or

$$4\cos 2\alpha = 3\sin 2\alpha$$

whence $\tan 2\alpha = 4/3$ or $2\alpha = 53°$ nearly.

In order to convert the implicit function of **No. 30** into an explicit function of x we express it as a quadratic equation in y:—

$$6y^2 - (x + 7)y - 2(x^2 - 1) = 0$$

of which the solution is

$$y = \frac{(x + 7) \pm (7x + 1)}{12}.$$

Taking in succession the *plus* and *minus* we have that the original function is equivalent to the two linear functions

$$2x - 3y + 2 = 0 \text{ and } x + 2y - 1 = 0.$$

CHAPTER XXXVII.

THE PROGRAMME OF EXERCISES LXVI-LXIX.

§ 1. *Mean Position* (Ex. LXVI).—Division A of this exercise may be regarded as supplementing, by the introduction of directed numbers, part of the argument of Ex. XXVI, D (Statistics). In that group of examples the student learnt to estimate the "mean deviation" of a series of measurements or other numbers from the "median" or middle term of the series. When the deviations of the individual terms of the series are measured, not by signless, but by directed, numbers their average may conveniently be called their "mean position". It is important to observe that whereas the mean deviation of the series will be different if measured from different terms—the mean deviation from the median being its least value—the mean position is independent of the point from which the individual deviations are estimated. This property is meant to be brought out in **Nos. 2, 3,** but is so essential that it should be emphasized by the teacher. The following may be given as a formal proof. Let the n numbers be, or be represented by, n points on a line, their distances from the point O, from which the deviations are to be reckoned, being $x_1, x_2, x_3, \ldots x_n$. In general some of these numbers will be positive and some negative. By definition the mean position of the numbers or terms with respect to O is

$$(x_1 + x_2 + x_3 + \ldots + x_n)/n \text{ or } (\Sigma x)/n$$

Now let the mean position be estimated with respect to a point O′ distant a from O. Then, whether a is positive or negative, the distances of the individual points from O′ will be $x_1 - a, x_2 - a, x_3 - a$, etc., and the mean position

$$(\Sigma x - na)/n = (\Sigma x)/n - a$$

Thus, though the mean position is represented by a different *number* it is still (or is still represented by) the same *point* as

before. For if (as in the note after No. 1) we put $\bar{x}$ for $(\Sigma x)/n$, the distance of the mean position from O, the point whose distance from O′ is $\bar{x} - a$ is obviously identical with it. The practical consequence is that, in finding mean positions or "centroids," we may count the individual deviations from any point or value which seems convenient.

These considerations show that the centroid or point of mean position may be regarded as essentially a statistical idea which has geometrical as well as arithmetical applications. As the reader is aware, it has also important applications in mechanics where it appears as (or, rather, underlies) the conception of "centre of mass" or "centre of gravity". The examples in divisions B and C do not go beyond the geometrical applications which may be regarded as of universal interest, but they will form a very useful introduction to certain parts of mechanics in the case of students who take up that study. The method followed is suggested immediately by the argument of ch. XXVII., A, and the examples may, as we have already said, be looked upon as illustrations of the power and usefulness of Wallis's "arithmetic of infinites".

In **Nos. 7-10** since there are n squares in the unit of length there are nkh in AB and $nkh \times x/h = nkx$ in a row distant x from O. Again, there will be nh rows, or $nh + 1$ if we count O itself as a row. In the pth row from O there will be $nkh \times p/nh = pk$ squares, the centres of which are at a distance p/n from the line through O since the distance between consecutive rows of centres is $1/n$. Thus the total distance of the centres of the squares in the pth row from the line through O is $pk \times p/n = p^2 \times k/n$. Since there are $(nh + 1)$ rows, the numbers of squares in which increase in A.P. from 0 to nkh, the total number of squares is $\frac{1}{2}nkh\,(nh + 1)$. To find the mean distance of all the points from the line through O we have then

$$\bar{x} = \frac{(0^2 + 1^2 + 2^2 + 3^2 + \ldots + n^2h^2)k/n}{\frac{1}{2}nkh(nh + 1)}$$

$$= \frac{0^2 + 1^2 + 2^2 + 3^2 + \ldots + (nh)^2}{(nh + 1)\,(nh)^2} \times 2h$$

$$= \left(\frac{1}{3} + \frac{1}{6nh}\right) \times 2h$$

by Wallis's theorem (p. 280). We conclude that when n is so large that the squares may be considered points the mean distance of the surface of the triangle from the line through O is $\frac{2}{3}h$.

No. 13 is to be done similarly, with the difference that we must here find an expression for the mean distance of the centres of a number of small cubes from the plane through the apex of the cone parallel to the base. Let h be the height of the cone and A the area of the base, and let the side of each cube measure $1/n$ of a unit so that n^2 of them would stand on a square unit. Then there are nh layers of cubes (or $nh + 1$ if the apex is counted as a layer), $n^2\mathrm{A}$ cubes in the bottom layer and $n^2\mathrm{A} \times p^2/(nh)^2 = p^2 \times \mathrm{A}/h^2$ in the pth layer from the apex. The distance of the centre of each cube in the pth layer from the plane through the apex is p/n and the total distance of all their centres is

$$\frac{p}{n} \times p^2 \times \frac{\mathrm{A}}{h^2} = p^3 \times \frac{\mathrm{A}}{nh^2}.$$

The total number of cubes in the whole cone will be

$$\frac{\mathrm{A}}{h^2}(0^2 + 1^2 + 2^2 + \ldots + n^2h^2) = \frac{\mathrm{A}}{h^2} \times \tfrac{1}{6}\, nh(nh + 1)\,(2nh + 1)$$

by the argument on p. 280. Also the total distance of the centres of all cubes from the plane through O will be

$$(0^3 + 1^3 + 2^3 + \ldots + n^3h^3) \cdot \frac{\mathrm{A}}{nh^2}.$$

The mean distance is therefore given by

$$\bar{x} = \frac{\left\{0^3 + 1^3 + 2^3 + \ldots + (nh)^3\right\} \dfrac{\mathrm{A}}{nh^2}}{\tfrac{1}{6}\, nh(nh + 1)\,(2nh + 1) \cdot \dfrac{\mathrm{A}}{h^2}}$$

$$= \frac{0^3 + 1^3 + 2^3 + \ldots + (nh)^3}{(nh + 1)\,(nh)^3} \times \frac{6nh^2}{2nh + 1}$$

$$= \left(\frac{1}{4} + \frac{1}{4nh}\right) \times \frac{3h}{1 + \dfrac{1}{2nh}}.$$

When n is so large that the cubes may be considered points making up the whole volume of the cone we have

$$\bar{x} = \tfrac{3}{4}h.$$

In **No. 15** the spherical shell is to be supposed divided into a number of zones or bands by equidistant planes parallel to the plane of section. It is supposed that the student knows that these bands all have the same area—namely that of the corresponding band on a cylinder circumscribing the ball with its axis perpendicular to the plane of section. [This proposition is proved in Part II, Ex. LXXXVI.] It follows that the mean position of the spherical shell is identical with that of the circumscribing cylinder of the same height—i.e. that it is on the axis of the shell mid-way between the plane of section and the remaining pole of the shell.

The examples in division C give apt and important illustrations of the power residing in the notion of mean position; they are not likely to cause difficulty.

§ 2. *Root-Mean-Square Deviation* (Ex. LXVII).—The reader is doubtless aware that in the analysis of many important physical phenomena (for example, those of rotation) it is necessary to take account of the *square* of the distance of each particle of a body from a certain point, axis or plane. Just as calculations concerning "centres of mass" are best regarded as special applications of the idea of mean position, so calculations concerning "moments of inertia" and similar physical entities are thought of most profitably as applications of another general statistical notion—that of "root-mean-square deviation". As will be shown in Part II, Section IX, this notion plays, in the general theory of statistics, a part which makes it a very important subject of study quite apart from its usefulness to the student of mechanics or physics. It is introduced here partly on account of this far-reaching significance and partly because it gives excellent opportunities for applying Wallis's Law to simple and interesting problems.

The definition of the root-mean-square measure of the dispersion or degree of "scatter" of a series of measurements is given in the note at the beginning of the exercise, together with two reasons for its superiority to the simpler "mean deviation" studied in Ex. XXVI, D. From the mathematical point of view the former is the more important, and may be expressed more fully as follows. The method of mean deviation is simple and effective when applied to straightforward arithmetical instances but leads to difficulties in an algebraic treatment because it ignores the signs of the deviations. By squaring the deviations we can at once avoid these difficulties

and retain the advantages of dealing with numbers which have all the same sign.[1] For this reason and others Prof. Karl Pearson has given to the root-mean-square measure the name of the "standard deviation" by which it is now generally known. The use of this term is reserved until the systematic investigation of statistical methods is undertaken in Section IX.

No. 3 is of great importance but is quite easy. If r is the distance of one of the n points from the origin we have

$$r^2 = x^2 + y^2$$

whence $$(\Sigma r^2)/n = (\Sigma x^2)/n + (\Sigma y^2)/n.$$

Representing the root-mean-square deviation of the points from the two axes and the origin respectively by the symbols D_x, D_y and D, this relation can be expressed in the form

$$D^2 = D_x^2 + D_y^2$$

whence we have

$$D = \sqrt{(D_x^2 + D_y^2)}$$

which is the required formula.

The solutions of Nos. 6-20 follow lines made familiar by Ex. LXVI. In **No. 6** the sum of the squares of the distances of the centres of the beads from the centre of the end one is

$$(0^2 + 1^2 + 2^2 + \ldots + m^2)\, d^2$$

Since their number is $(m + 1)$ the mean-square distance is

$$D^2 = \frac{(0^2 + 1^2 + 2^2 + \ldots + m^2)d^2}{m + 1}$$

$$= \frac{(0^2 + 1^2 + 2^2 + \ldots + m^2)d^2m^2}{(m + 1)m^2}$$

$$= \left(\frac{1}{3} + \frac{1}{6m}\right)d^2m^2.$$

Noting that $dm = l$ we conclude that

$$D = l\sqrt{\left(\frac{1}{3} + \frac{1}{6m}\right)}.$$

To obtain the answer in **No. 7** we must suppose m so great that its reciprocal may be neglected; hence $D = l/\sqrt{3}$.

In **No. 8** we assume that there are $2m + 1$ beads whose centres are d cms. apart and find the mean-square distance of the centres from that of the middle one by the method just illustrated. Then

[1] See Yule, *Theory of Statistics*, pp. 134, 146.

$$D^2 = \frac{2(0^2 + 1^2 + 2^2 + \ldots + m^2)d^2}{2m + 1}$$
$$= \frac{0^2 + 1^2 + 2^2 + \ldots + m^2}{(m + 1)m^2} \cdot \frac{2m^2d^2(m + 1)}{2m + 1}$$
$$= \left(\frac{1}{3} + \frac{1}{6m}\right) \cdot \left(1 - \frac{1}{2m + 1}\right) \cdot \frac{l^2}{4}$$

since in this case $md = l/2$. Making m an exceedingly large number, we have

$$D^2 = \frac{l^2}{12}$$

or $D = l/2\sqrt{3}$. The same result can be obtained in a simpler manner by applying No. 6 to the two halves of the line.

The result of No. 8 is easily seen to apply also to **No. 9**, the beads being replaced by thin rods parallel to CC′. It is less obvious that the two problems of **No. 10** have the same solutions as Nos. 7 and 8. Remembering, however, that a spherical surface may be supposed divided up into narrow zones of equal depth whose areas will be equal to those of corresponding belts of the circumscribed cylinder, we see that the root-mean-square distances of the points of the bubble from the two planes mentioned in the question will be the same as the corresponding distances in the case of the cylinder. But it is obvious that a hollow cylinder may be treated in the same way as a line or a rectangle.

Nos. 11 and **15** require practically the same argument. In the latter case suppose that the distances between the circumferences of consecutive circles is d and that there are p points or beads on the circumference of the circle of radius d. Then the circle whose radius is (say) sd will have sp beads. The sum of the squares of their distances from the centre will be

$$(sd)^2 \times sp = s^3 \times pd^2$$

Thus the sum of the squares of the distances of all the beads from the centre is

$$(0^3 + 1^3 + 2^3 + \ldots + m^3)pd^2$$

The total number of beads is

$$(0 + 1 + 2 + 3 + \ldots + m)p = \tfrac{1}{2}m(m + 1)p$$

We have, therefore, for the mean-square distance

$$D^2 = \frac{(0^3 + 1^3 + 2^3 + \ldots + m^3)pd^2}{\frac{1}{2}m(m+1)p}$$

$$= \frac{0^3 + 1^3 + 2^3 + \ldots + m^3}{(m+1)m^3} \cdot 2m^2d^2$$

$$= 2r^2 \cdot \left(\frac{1}{4} + \frac{1}{4m}\right)$$

$$= r^2\left(\frac{1}{2} + \frac{1}{2m}\right)$$

since $r = md$. Hence

$$D = r\sqrt{\left(\frac{1}{2} + \frac{1}{2m}\right)}$$

When m is supposed so large that $1/2m$ can be suppressed we have the answer to **No. 16.**

§ 3. *The Binomial Theorem* (Ex. LXVIII).—In Ex. XXXI the student learnt how to obtain the "expansion" of $(1 + a)^{n+1}$ when the expansion of $(1 + a)^n$ is given, n being any positive integer. As he then saw, the method is an ancient one, going back at least to the German algebraist Stifel (*c.* 1544). In 1665 Newton, by one of those "accidents" which are reserved for men of genius, was led to make two observations which converted a cumbrous rule of limited application into a general theorem of fundamental importance. The first was that the coefficients in the expansion of $(1 + a)^n$ can be calculated directly for a given value of n without reference to the coefficients belonging to any other value; the second that n, in his expansion-formula, need be neither integral nor positive. The discovery was made in the course of an attempt to obtain a series for π by a method suggested, but not completed, by Wallis.[1] But although we are justified in commemorating the great Englishman's connexion with the binomial theorem by the term "Newtonian coefficients," it should be remembered that he never succeeded in giving a general proof of its validity.

Tradition has assigned to the binomial theorem an important place in the school course but it has done little to justify it. The treatment of Ex. LXVIII is intended to show that there are reasons for studying the expansion perhaps more substantial than its "elegance," and more immediately con-

[1] For details see the author's article in the *Mathematical Gazette* for January, 1911.

vincing than its usefulness in the advanced parts of the subject. It will be seen that the theorem is presented as an "approximation-formula" which is a generalization of the simple approximation-formulæ of Exs. IX-XI, etc., and is capable of numerous important applications. The starting-point of the discussion—the approximate calculation of compound interest—was suggested by a famous address of the engineer George Bidder to the Institute of Civil Engineers.[1]

The exposition given in the notes and examples of the exercise is sufficiently full to make further explanation unnecessary. The teacher should, however, take care to emphasize the argument of **Nos. 16-19**. The aim here is to find a means of estimating the exactness of the approximation yielded by a given number of terms of the expansion of $(1 - i)^{-n}$ where n is a positive integer. It is shown to be impracticable to calculate exactly the value of the "complement" as was done in Ex. XXXV, but it is found to be possible, by an application of the theory of geometric series, to name, in certain cases, a number which is certainly *greater* than the complement. Whenever, therefore, it is possible to use the expansion of $(1 - i)^{-n}$ or $(1 + i)^{-n}$ as an approximation-formula it is also possible to find an outside estimate of the error involved in limiting it to a given number of terms.

The argument of division B may be regarded as a demonstration of the binomial theorem for a negative integral exponent. A demonstration in the case of a fractional exponent is outside the scope of an elementary work and the student must, like Newton himself, be contented with a verification in certain instances. A simple and convincing form of verification, based upon the properties of the exponential curve, is suggested in division C. The examples given to illustrate the practical value of this case of the binomial expansion regarded as an approximation-formula contain no difficulties calling for comment.

[1] Reported in the *Minutes of Proceedings* of the Institute for 1856. An interesting account of the paper, with extensive quotations, is given by Mr. Branford in his *Study of Mathematical Education*, ch. VII., but the reader must consult the original for Bidder's method of solving problems of compound interest. The author owes his first acquaintance with the paper to Mr. Branford's citations.

The short division D is intended to suggest a means of calculating the value of "e," a number which has been estimated hitherto only by a rather precarious graphical method. The student is to see that as n increases the coefficients in the expansion of $(1 + 1/n)^n$ approach constantly nearer to those of the formula given in **No. 34**. This formula may, therefore, be adopted with fair confidence as the one to be used in calculating the value of e to any required degree of accuracy. The teacher should, of course, point out that the argument must not be regarded as a sufficient demonstration. For this the student must wait till Section VIII is reached.

§ 4. *The Generalization of Wallis's Law* (Ex. LXIX).—With the binomial theorem in our hands it is easy to show that Wallis's Law holds good for all values of the exponent of x. So far as positive and negative integral values are concerned the following proof may be regarded as satisfactory. Let $y = ax^m$ where m is either a positive integer n (**No. 1**) or a negative integer $-n$ (**No. 2**). Then if (as in Ex. XLVIII) a series of ordinates be drawn to the graph of the function at equal distances h the first difference of y will have the value

$$a(x+h)^m - ax^m = a\left\{mx^{m-1}h + \frac{m(m-1)}{2!}x^{m-2}.h^2 + \frac{m(m-1)(m-2)}{3!}x^{m-3}.h^3 + \ldots\right\}$$

$$= amx^{m-1}h + amh^2\left\{\frac{m-1}{2!}x^{m-2} + \frac{(m-1)(m-2)}{3!}x^{m-3}.h + \ldots\right\} \ldots \text{A.}$$

Now if m is positive there is a definite number of terms in the bracket in A and their sum will therefore have a definite value when numbers are substituted for the symbols. Call this sum S. Then the ratio to the term $amx^{m-1}h$ of the part of the first difference which follows it is

$$\frac{amh^2 . \text{S}}{amx^{m-1}h} = \frac{h\text{S}}{x^{m-1}} \quad . \quad . \quad . \quad \text{B.}$$

No matter how small x may be, h can be taken so much smaller that the ratio may be reduced below any specifiable number. In other words, when h is sufficiently small the value of the first difference of y becomes equal to the term $amx^{m-1}h$ within c per cent, c being as small a number as any

one pleases to name. If m is negative there is no definite number of terms within the bracket in A and they have, therefore, no definite sum. But the argument of Ex. LXVIII, B, showed that if h/x is numerically small enough it is at least possible to specify a number S which is greater than the sum of any conceivable number of terms within the bracket. Supposing S in B to be this number the rest of the argument follows as before. But if all terms in A after the first may be neglected h is the differential of x and the first difference is the differential of y. We have, therefore, for all integral values of m the differential formula

$$\frac{\delta y}{\delta x} = amx^{m-1}.$$

When m is fractional the proof ceases to be so satisfactory for we have not *demonstrated* the binomial theorem in this case —we have only made it very probable. The note before **No. 7** gives an argument in which an attempt is made to evade this difficulty. It is, perhaps, sufficiently good for this stage of the student's development but its theoretical imperfection should not be overlooked. Assuming the foregoing proofs that R and R′ can be reduced below any given percentage of px and qx' by taking x and x' small enough, it follows, strictly speaking, only that

$$x' = \frac{p}{q}x$$

to an indefinitely close degree of approximation. It is also assumed that as x is reduced x' may be brought below unity.

The notes and examples of divisions B and C show how differential formulæ based upon Wallis's Law can be used to solve problems of area-evaluation and the determination of the distance traversed by a point moving under specified conditions. The arguments require careful consideration but are set out very fully in the text. The teacher will doubtless remember in connexion with them what was said in ch. XXIV., § 7, about the relation of these reasonings to those involving the notion of a "limit".

The exercise concludes with a few examples (division D) illustrating the application of different differential formulæ to certain geometrical problems. These are not to be considered of particular importance, but generally prove interesting to the more able young mathematicians in the class.

It may be useful to add the solutions of a few typical problems taken from the various divisions of the exercise.

In the second part of **No. 9** it must be remembered that h is always the *increment* of x. Thus when the *plus* before the x becomes *minus* the sign before the h must also be changed. It follows that the differential formula of $y = (a - x)^m$ is

$$\frac{\delta y}{\delta x} = -m(a - x)^{m-1}.$$

The results both in **No. 9** and in **No. 10** are proved very simply by a graphic method which should be demonstrated to the class. By Wallis's Law the gradient of the tangent at a given point of the curve $y = x^m$ is mx^{m-1}, x being the abscissa of the point. Move the curve a distance a to the left. Then the number previously symbolized by x must now be symbolized by $a + x$. That is, $y = (a + x)^m$ and the gradient is $m(a + x)^{m-1}$. Again, it is easy to show from first principles that if $y = (-x)^m$ the gradient is $-m(-x)^{m-1}$. Hence by the previous argument the gradient of $y = (a - x)^m$ is $-m(a - x)^{m-1}$.

This result is applied in **No. 13 (iii).** Since $y = \delta A/\delta x$ we have

$$\begin{aligned}\frac{\delta A}{\delta x} &= (1 - 3x)^{-4} \\ &= \{3(\tfrac{1}{3} - x)\}^{-4} \\ &= 3^{-4} . (\tfrac{1}{3} - x)^{-4}\end{aligned}$$

whence

$$\begin{aligned}A &= 3^{-4} . \tfrac{1}{3}(\tfrac{1}{3} - x)^{-3} \\ &= \tfrac{1}{9}(1 - 3x)^{-3}\end{aligned}$$

The accuracy of this solution is seen when the result reached in No. 9 is applied to the formula $y = \frac{1}{2\,4\,3}(\frac{1}{3} - x)^{-3}$.

The fact stated in **No. 14** is of much theoretical and historical importance.[1] Wallis's Law breaks down when $m = 0$ because in this case the formula for the primitive, x^m/m, takes the meaningless form 1/0.

[1] See Part II, Ex. LXXXIV, G. For the history see the author's article in the *Mathematical Gazette*, for December, 1910.

No. 16 (iii) gives another illustration of the method of No. 9. We have

$$\frac{\delta s}{\delta t} = 100(1 + 5t)^{-\frac{1}{2}}$$

$$= \frac{100}{\sqrt{5}} \cdot (0{\cdot}2 + t)^{-\frac{1}{2}}$$

whence

$$s = \frac{100}{\sqrt{5}} \cdot 2(0{\cdot}2 + t)^{\frac{1}{2}} + a$$

$$= 40(1 + 5t)^{\frac{1}{2}} + a$$

a being the undetermined constant. To find the value of a we note that, when $t = 0$, $s = 0$. Hence $a = -40$. Thus we have finally

$$s = 40\{(1 + 5t)^{\frac{1}{2}} - 1\}.$$

In **No. 18 (iv)** we have

$$\frac{\delta^2 s}{\delta t^2} = -4(1 + 2t)^{-3}$$

$$= -\tfrac{1}{2}(\tfrac{1}{2} + t)^{-3}$$

whence

$$\frac{\delta s}{\delta t} = \tfrac{1}{4}(\tfrac{1}{2} + t)^{-2} + a$$

$$= (1 + 2t)^{-2} - 1$$

since $\delta s/\delta x = 0$ when $t = 0$. Again in **No. 19 (iv)** we have

$$\frac{\delta^2 s}{\delta t^2} = \tfrac{1}{4}(\tfrac{1}{2} + t)^{-2} - 1$$

whence

$$s = -\tfrac{1}{4}(\tfrac{1}{2} + t)^{-1} - t + b$$

$$= -\tfrac{1}{2}(1 + 2t)^{-1} - t + \tfrac{1}{2}$$

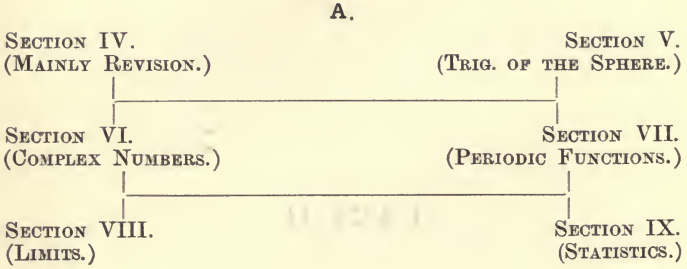
since, when $t = 0$, $s = 0$.

PART II.

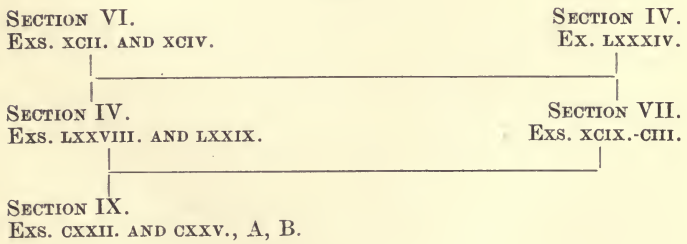
ALTERNATIVE SCHEMES OF STUDY.

(Cf. pp. 382, 398.)

A.



B.



CHAPTER XXXVIII.

INTRODUCTION TO PART II.

§ 1. *The Scope of Part II.*—The aim of Part I is to set forth a course which, though liberal in range, is confined to fundamentals. That is to say, it contains no subject of which a student leaving a secondary school should be permitted to be ignorant, and treats no topic in a way relevant only to the needs of the specialist. The scope of Part II goes beyond fundamentals, at least in the sense that none of its contents can be said to be an indispensable part of a universal minimum course. On the other hand, it is still confined to them in the sense that the choice of subject-matter and methods of treatment are based upon the ideas set forth in ch. II., § 1, that is, upon the conception of mathematics as a cultural rather than a technical subject. Text-books on "higher algebra" are, as a rule, written professedly for the aspirant to mathematical honours at the University; they aim, therefore, at a rather advanced development of technique within a somewhat limited field. The question whether it would not be better in every case to reserve studies of this kind until the student actually reaches the University or higher technical institution is important but cannot be argued here. It is, however, proper to urge that they do not offer the most suitable course of instruction for the general body of students. For the student who is to be a teacher or an engineer, or to engage in higher industrial or administrative work, as well as for the student who is continuing his mathematical studies as part of a general education, the best course would seem to be one which sets in clear relief the central aims and most vital notions of the main branches of mathematics, supplements exposition with sufficient practical exercise to give the student a real training and the sense of mastery that comes with training, and, in particular, illustrates vividly the essen-

tial part which mathematics plays in so many departments of modern life and activity.

If there is room for individual choice and preference in the details of the minimum course no specific programme for a further course can claim to be much more than an expression of the views and, perhaps, the temperament of the proposer. Moreover, the variety in the conditions under which mathematical teaching is carried on in this country makes the actual adoption of a uniform programme impossible even if it were desirable. For these reasons *Exercises*, Part II, is to be regarded simply as setting forth the full range of topics which the author would himself teach under favourable conditions to pupils of fair ability. He recognizes that other teachers who use the book will, for one reason or another, utilize only a selection from its contents. To facilitate selection the various sections have been made as independent of one another as the nature of the subject-matter permits, and the exercises of each section have been divided into those which are essential to the development of the subject and those which may be omitted by the student for whom a shorter course is prescribed.

The full course consists of six sections, numbered IV-IX in continuation of those of Part I. Algebraic theory and practical applications appear side by side in each; but it is useful to note that the development of theory is predominant in three of the sections (IV, VI, VIII) while practical applications form the central interest of the sections which alternate with them. Scheme A on p. 380 shows the order in which the sections are intended to be studied when the full course is taken. The idea underlying the arrangement is that a "theoretical" and a "practical" section shall always be studied at the same time. Even if a selection is made from the full course it would probably be best in most cases to preserve this feature of the scheme. Scheme B suggests a minimum course in which at least those subjects will be studied which are commonly required in examinations, such as the Intermediate B.A. and B.Sc. Examination of the University of London, the "Advanced" paper in the Board of Education's Examination of students in Training Colleges, and several professional and technical examinations of a similar standard (see pp. 60 and 398). The aim and scope of the several sections will now be indicated.

§ 2. Section IV. *Mainly Revision.*—The title of this section indicates that it looks backwards towards Part I rather than forwards to the rest of Part II. As is explained in the Introduction (*Exercises*, II, pp. 3, 4), its aim is not only to consolidate and extend the knowledge gained in the earlier part of the course but also to make explicit, and to bring into scientific order, ideas about the nature of numbers, of arithmetical processes, and of algebraic symbolism which are in one way or another involved even in the first steps of the student's mathematical progress, though he has not hitherto made them the subject of systematic reflexion. The exercises fall into three well-marked groups. Exs. LXX-LXXIV deal with the nature of numbers, Exs. LXXV-LXXVIII with the idea of a function and the properties of algebraic symbolism, Exs. LXXIX-LXXXIII with certain more advanced applications of the ideas of Sections II and III. The supplementary Ex. LXXXIV consists chiefly of revision papers upon the subject-matter of Part I.

Some considerations about the properties of numbers and arithmetical operations are a common feature of elementary text-books on algebra and are usually summarized in the "laws" of commutation, of association, etc. Too often these appear at a point in the course where the pupil is not prepared to appreciate their significance—namely, at the beginning. General pedagogical principles suggest that reflexions upon the assumptions involved in a science should come after the student has made some progress in the science by the light of "intuition" and common sense. The postponement of these discussions about the foundations of algebra to the place here assigned to them not only gives them a better chance of being appreciated but also makes it possible to expand them into a simple exposition of the main results of the modern theory of the nature of numbers.[1] Numbers and their relations to the great twin entities, space and time, have always exercised an attraction upon thoughtful minds. Philosophies have been based upon the properties of numbers; religions upon their power and mystery. In particular, the seeming contradictions involved in the concept of "infinity"

[1] This is, of course, an entirely different subject from the "theory of numbers" which is one of the traditional topics of a course of "higher algebra".

which provoked the acutest of the Greeks to "arguments, all immeasurably subtle and profound" [1] continue to "tease out of thought" ordinary folks as well as poets. But in spite of this universal interest in the subject the philosophy of number could not, until comparatively recently, be considered a suitable topic for elementary instruction. The labours of modern philosophical mathematicians—especially the "almost unexampled lucidity" of Cantor [2]—have, however, made it possible to present and solve the main problems concerning the nature of numbers in a way both interesting and profitable to the general student. Such a presentation is attempted in Exs. LXX-LXXIV. Current opinion does not as yet demand that these problems shall be discussed in schools, but they have a character which appeals to the expanding interests of the adolescent and are of profound importance outside the boundaries of technical mathematics. The teacher is recommended, therefore, to find a place for them in his course. If he needs further persuasion to do so he is invited to observe that without some consideration of the relations between numbers and continuous quantities—such as lengths and periods of time—many important conclusions and practices in elementary algebra rest upon a very unsatisfactory logical basis.

Exs. LXXV-LXXVIII stand less in need of defence. Exs. LXXV-LXXVI are intended to revise and give further illustrations of ideas about the nature of functions of a single variable which have already been acquired in Part I. The essentials of the ideas connoted by the terms "indeterminate value" and "singular points" find their place here. In Ex. LXXVI an inquiry into the properties of a few functions of two variables is made the occasion for extending the method of rectangular coordinates to the analysis and description of curved surfaces. The investigation is undertaken in the spirit of ch. IV., § 9; that is, the surfaces are treated as tri-dimensional graphs to be studied not so much for their own sake as for the light they throw upon the properties of the functions of which they are the spatial expression. The exercise should, however, prove a useful introduction to tri-

[1] Russell, *Prin. of Math.*, p. 347. The reference is to the paradoxes of Zeno. (See Ex. LXXII, No. 4.)

[2] Russell, *op. cit.*, p. 353. The author's acquaintance with the subject is derived chiefly from Mr. Russell's masterly pages.

dimensional coordinate geometry in the proper sense of the term. The last exercise of the group is given to a brief review of the principles which govern the development of algebraic symbolism in all its stages. The fact that it uses the exponential notation as the chief illustration of its theme makes it a suitable link between the two "theoretical" subsections and the "practical" subsection which follows.

The exercises of this last group (Exs. LXXIX-LXXXIII) are all concerned with the law of "growth with constant growth-factor" and the exponential function which that form of growth exemplifies. Exs. LXXIX-LXXXI are frankly informative, being intended to exhibit the essentials of the theory of public loans (and similar financial arrangements) and the principles underlying the practice of life assurance. It should be unnecessary to demonstrate the immense importance of these parts of the "mathematics of citizenship"; the common neglect of them in schools is perhaps due only to an exaggerated estimate of their difficulty. This remark applies especially to the theory of life assurance which is generally supposed to require previous acquaintance with the theory of probability. It will be seen that the typical problems of the life office can be solved without the introduction either of the term "probability" or of the ideas which the term is thought to connote. In Ex. LXXXII the student turns from these topics to a more formal study of the exponential function and curve, theoretical conclusions being illustrated by references to simple physical phenomena such as those exhibited by cooling bodies. In Ex. LXXXIII he considers the differential formulæ of the exponential and logarithmic functions —a subject of great historical and practical importance.

Ex. LXXXIV is divided into six parts. The first five are Revision Papers and consist almost entirely of straightforward exercises on the subject-matter of Part I. The exception to this statement is that in a few examples in division E the simple theory of the decimal notation outlined in Ex. XXXI is expanded into a brief treatment of "scales of notation" in general. In the last division (F) the student follows the steps that led Wallis, Mercator, and others to discover the expansions by which logarithms may most conveniently be calculated. With the exception of division F the whole of this exercise is independent of the rest of the section and may be given as tests at any convenient points.

§ 3. Section V. *The Trigonometry of the Sphere.*—Perhaps the most serious defect in the English curriculum in mathematics is its neglect of what continental writers call "stereometry," that is, the elementary study of three-dimensional space by the methods of geometry, algebra and trigonometry. At various points of Part I examples have been introduced with the object of freeing the pupil from intellectual confinement to the plane; these have been followed in Section IV by an exercise (LXXVII) dealing specifically with graphic surfaces and the solid forms they enclose. Section V is devoted entirely to a systematic though elementary treatment of stereometry in one of its most important branches—namely, the trigonometry of the spherical surface. We have seen (ch. II., § 3) that the problems of spherical trigonometry have hitherto been excluded from the schoolroom—in spite of their intrinsic interest and great practical importance—mainly because they have been regarded as a distinct "subject" whose place in the logical hierarchy comes after plane trigonometry. The obvious way to escape from the consequences of this pedagogical error is to select a group of spherical problems of general rather than technical interest and to investigate trigonometrical methods of dealing with them—the subject being developed no farther and with no more formality than the solution of these problems itself requires. The increasing attention given to the study of geography in schools suggests as the appropriate starting-point a simple investigation of the theory of map-projections. Some consideration of this subject must necessarily find a place in any rational course of instruction in geography; thus to begin here is to build upon ideas with which the student has already a certain familiarity. Moreover, some of the cartographic projections which are found in every school atlas are based upon principles of the highest interest and beauty from the purely mathematical standpoint. Lastly, the study of these concrete instances leads simply and naturally to the general notion of a "transformation," which is one of the most significant and vital of mathematical ideas.

The first thing which a student of this subject learns is that the points of a spherical surface can be represented upon a plane in an endless number of ways, and that, in practice, the choice of a "projection" is always determined by the particular purpose it is to subserve. One purpose of special

interest and importance is to assist the sailor in the task of navigating his ship from port to port. The study of the charts constructed for this end leads naturally to an investigation of the trigonometrical formulæ of which the navigator has to make daily use. In this way the student reaches the problems of spherical trigonometry in the narrower sense of the term.

The topics indicated in the two preceding paragraphs form the subject-matter of the first group of exercises in Section V —namely, Exs. LXXXV-LXXXIX. Of these the first four are given to the theory of the chief systems of map-projection and to the graphical solution of the geographical problems associated with each. The modest outfit of trigonometrical formulæ which suffices to solve these same problems by calculation is all developed in the last exercise of the group (Ex. LXXXIX).

The doctrine of the sphere learnt in the first group is applied in another single exercise (XC) to a series of simple astronomical problems. This part of the work has an obvious connexion with the study of navigation but is justified less by that circumstance than by its wider connexion with geographical theory and by its intrinsic interest. *Some* study of astronomical phenomena is a necessary feature of every course in geography. As in the parallel case of map-projections, to make precise, to develop, and to apply the ideas thus acquired by the pupil is at once the mathematical teacher's plain duty and his valuable opportunity. Apart from these utilitarian considerations, the neglect of the most ancient branch of "Nature study" is an anomaly in our school curriculum which ought no longer to be tolerated. If a few lessons, desirable in themselves from the standpoint of mathematical instruction, have the further happy result of helping to remove that anomaly the case for including them in any programme of moderately extensive scope is extremely strong.

It should be clearly understood that this is not an argument for the systematic study of astronomy in schools under the name of "applied mathematics". It is urged merely that certain lessons in mathematics should have the incidental result of teaching boys and girls to "lift up their eyes to the heavens" in a spirit of intelligent inquiry. In other words, the astronomical problems to be considered in the mathe-

matical classroom must be limited to those whose solution is simply a question of the straightforward mathematical statement of facts accessible to ordinary observation. The constitution of the solar system or the connexion between orbital motion and "central forces," though excellent subjects for special study in school, would be out of place in the general mathematics course. After the exclusion of technical matters of this kind there is still an abundance of interesting and important questions which our principle readily admits. They include (i) the problem of fixing the positions of the stars by co-ordinates, (ii) the related question of the diurnal revolution of the heavens, (iii) the daily movements of the sun and moon, (iv) the calculation of times of rising and setting, etc., (v) the nautical problems of determining latitude and longitude, and, lastly, (vi) the fascinating subject of dialling.

In the treatment of all these questions in Ex. XC the facts are stated as they appear to direct observation. There is no mention of the earth's daily rotation upon its axis or of its annual revolution around the sun. These are *hypotheses* invented to account for the facts of observation and belong, therefore, to a technical study of astronomy. We are limited here to a precise statement of the face-value of the facts which we observe. According to that face-value the stars move round the sky daily and the sun and moon move among them. Any attempt to give a theory of stellar movements must begin with an exact determination of the facts as they appear; so that the method pursued in Ex. XC is one which lays the foundation indispensably necessary for any successful systematic study of astronomy.

It cannot be assumed that the average student has a knowledge of astronomical facts equal in scope to the geographical knowledge assumed in the treatment of problems of navigation. It has been necessary, therefore, to summarize in some detail what direct observation has to tell about stellar and solar movements. A probable cause of the widespread neglect of simple astronomical knowledge is the abstract and technicalized form in which the systematic text-book presents it. Happily in many schools this neglect is passing away, and simple studies of solar shadows, etc., are being added to the now customary programme of meteorological observations. Experience shows that there is no difficulty in establishing at first hand, and without any instrumental

means more recondite than an upright pin and a home-made sundial, all the facts of the diurnal and annual movements of the sun and stars upon which rest the theories of day and night and of the seasons. Experience also shows that there are few branches of study which engage more readily the interests of boys and girls, and none which gives such direct aid to mathematical progress.[1] It is not too much to say that the general adoption of simple observations of this kind, followed up by elementary discussions in the classroom, would alone suffice to put an end to that neglect of tridimensional geometry which we have signalized as perhaps the most serious deficiency in current mathematical instruction.

It is not possible to describe here the practical exercises which would satisfy the conditions mentioned above. The most important are probably suggested with sufficient directness in the exposition of Ex. XC, and the teacher will have no difficulty in giving practical effect to the suggestions —particularly if he make common cause with a colleague who teaches geography. The globe with the blackboard surface, recommended for use in connexion with the exercises of Section V, can easily be converted into an astronomical globe by the addition of a cardboard horizon, and is then much more useful than the globe of the makers, which contains too much detail for the teacher's purpose. If a copy of that excellent old text-book, Keith's *Use of the Globes*, can be obtained, it will be found to give all the information which can be needed. It must, however, be emphasized that the globe should be employed as a means of summarizing and explaining observations which have actually been made, and that to use it as a substitute for first-hand observation is a thoroughly unpedagogical practice.

The supplementary examples of this section (Ex. XCI) deal partly with practical problems of greater difficulty and less fundamental importance than those of the former exercises. In addition they develop a little further some of the theoretical consequences of the earlier arguments. Thus in division B the notion of a projection is generalized into the idea of point-to-point correspondence which may subsist either between lines, surfaces, or volumes. The argument here paves

[1] The author owes to a friend who speaks with authority the observation that vitality in mathematical studies and interest in astronomy have, in the past, constantly risen and fallen together.

the way for the study of functions of a complex variable which is to be taken up in Section VI. Again, in divisions A and D the doctrine of spherical triangles receives additions, some of which (e.g. the study of "spherical excess") are of almost purely theoretical interest.

§ 4. Section VI. *Complex Numbers.*—In ch. xxiv., § 4, it was stated that the "imaginary roots" of quadratic equations would be excluded from consideration in Part I to be discussed later as part of a systematic doctrine of "complex numbers". At first sight this procedure may seem in contradiction with the principle asserted with so much emphasis in the preceding article and elsewhere. If it is an error to hold back the methods of the calculus or of spherical trigonometry until they can be taught as self-contained subjects, how can it be right to pursue the opposite policy in the case of "imaginaries"? The answer to this question has already been indicated on p. 239. It is true that the doctrine of complex numbers originated in the investigation of the roots of equations. The roots which the earlier algebraists (like the students of our Part I) rejected as "impossibles" were recognized by Girard (1629) as specific solutions, and were formally distinguished as "imaginary," in opposition to "real," roots by Descartes in his *Géométrie* (1637). Nevertheless, there are three good reasons for departing at this point from the historical track. In the first place the admission of "imaginary" roots would blur the clearness of the connexion between the quadratic equation and the parabolic curve. Secondly, "imaginary" roots are needed only in order that every quadratic may be regarded as soluble; and logical completeness of this kind makes but a feeble appeal to the immature mathematician. Thirdly, although the use of "imaginary" numbers during the seventeenth and eighteenth centuries produced some notable and fruitful results (for example, De Moivre's Theorem, 1730), yet they were, and continued to be, more or less of a mystery until Argand and Gauss, at the beginning of the nineteenth century, reached the views of their nature which are set forth in Section VI. There seems, therefore, to be abundant reason to withhold all consideration of the subject until the pupil is in a position to study those views with profit.

A brief commentary, historical and philosophical, upon the exposition of Section VI will be found in ch. xliv. It will be

sufficient here to indicate the ground covered by the exercises. Exs. XCII, XCIII introduce the conception of $a + ib$ (where a and b are any directed numbers) as a "complex number" which can be used as a symbol either for the point whose rectangular co-ordinates are a and b or for the straight line joining this point to the origin. They also inquire in what sense the operations of addition and multiplication can be applied to these complex numbers. The most important single result of this investigation is the discovery that the manipulation of expressions of the form $a + ib$ is immensely simplified by treating the symbol i as if it were a number whose square is -1. The first fruit of this momentous discovery is De Moivre's Theorem.

In Exs. XCIV and XCV the student sees that the adoption of complex numbers leads to most important extensions of the results of elementary algebra. The first of these (Ex. XCIV) consists in the idea that an equation which has no root in the simple arithmetical sense may yet be satisfied by "complex" roots, and that, in other cases, ordinary or "real" roots may be supplemented by complex roots. The second (Ex. XCV) is the discovery that in the exponential function $y = ar^x$, where r has been regarded hitherto as necessarily non-directed or positive (ch. xxx., § 5), the introduction of complex values of y enables us to remove the restriction and to consider r as capable of any numerical value.

The force of these extensions may be expressed alternatively as follows. Let $y = f(x)$ be any function. Hitherto it has been taken for granted that x and y represent numbers taken from the complete one-dimensioned scale which runs from $-\infty$ through zero to $+\infty$. We have learnt that a given function will not, in general, assign a value to y for every value of x or a value to x for every value of y, but that it connects the values of the variables only over a certain field. We have learnt also that it is always possible to represent this field of the function by a graph drawn upon a plane. In Ex. XCIV we discover that, in some cases, the "real" values of y which are not given by any "real" value of x yet correspond to "complex" values of that variable, and that the field of the function as enlarged by the admission of these complex values of x can be represented by a three-dimensioned graph. Similarly, it is found in Ex. XCV that, in the case of other functions, where some "real" values of x have no

"real" values of y corresponding to them, the defect may be made good by taking account of "complex" values of y, and that the field of the function, thus enlarged, can again be represented by a graphic line drawn in three-dimensional space. The obvious completion of these ideas is the notion that $y = f(x)$ may represent a form of connexion in which both x and y may simultaneously have "complex" values. The pursuit of this notion is begun in Ex. XCVI where it is applied to very simple cases. It is observed that we now have to do with two variables, say $x = u + iv$ and $y = \mathrm{U} + i\mathrm{V}$, each of which is of two dimensions, so that three-dimensional space no longer suffices for the graphic representation of the field of the function. Following the example of Riemann we turn to the projective method of Section V as the simplest way of meeting the new need. Armed with this method we attack, in Ex. XCVII, the especially important case of the logarithmic function. The supplementary examples of Ex. XCVIII derive from this investigation the doctrine of the "exponential values" of the sine and cosine and apply it to the discussion of the sine and cosine functions when the independent variable is complex.

§ 5. Section VII. *Periodic Functions.*—The central purpose of Section VII is the development of the idea of periodic functions of a variable. A function such as

$$y = ax^2 + bx + c \text{ or } y = a + \sqrt{bx + c}$$

is non-periodic. That is to say, although different values of x may in some cases be associated with the same value of y, or different values of y with the same value of x, yet there is no regular repetition of a cycle of values of the one variable as the other progresses through its scale of possible values. The typical feature of a periodic function is, on the other hand, that it does exhibit such a succession of identical cycles. Periodicity in physical phenomena is among the most familiar of our experiences—"seed time and harvest, and cold and heat, and summer and winter, and day and night" being only the most conspicuous instances. Our task in this section is to find an algebraic language which shall be appropriate to the analysis and description of such phenomena and the numerical relations which they exemplify, just as the symbolism studied in previous sections is appropriate to the analysis and description of phenomena which do not exhibit periodicity.

The elements of such a language are readily obtained. We have already found, in connexion with problems of mensuration and surveying (Pt. I, Exs. XL and XLI), the need of extending the idea of the sine and cosine of an angle to include "angles" measured up to 360°; and this extension has made us familiar with the repetition of the series of values presented by the sines and cosines of angles between 0° and 90°. Our present task merely requires three further steps in the same direction: (i) We must reach the idea of an angle as a variable capable of an endless series of values, negative as well as positive. (ii) To each of these values we must assign a sine and cosine in such a way that all angles which differ in value by a certain definite amount shall have the same sine and cosine. (iii) Finally, just as in Section II notions of the sine and cosine were made independent of their original association with the triangle, so now they must be made independent even of their associations with angles. When we have reached this point the symbols $\sin x$ and $\cos x$ will mean simply numbers associated with a number x in accordance with certain definite laws, and it will be no longer necessary to regard x in the symbolism $\sin x$ as an angle any more than it is necessary to regard x in the symbolism x^2 as a length. It may stand for any variable which has another variable depending upon it in accordance with the special law of the function.

Thus conceived $\sin x$ and $\cos x$ are our first instances of a periodic function. The next step is to find that by combining sines and cosines we obtain formulæ which describe periodic relations of indefinite complexity. From this discovery it is an easy passage to the converse idea that any periodic function may be expressed by a formula built up of sines and cosines upon the simple plan first laid down by the great French mathematician, Fourier.

It is important that the teacher should keep closely before him the main argument of the section. The incidental results to which the argument leads have, however, much substantive importance and should be mentioned. Some are particular methods, such as the measurement of angles in radians, some are formulæ of wide applicability, such as the differential formulæ for the sine and cosine, some are important expansions, such as Gregory's series for π. The most considerable, however, is the idea of the hyperbolic

functions. The treatment of this topic will, it is hoped, be found simple enough to be followed without difficulty by the average student. The main idea to be emphasized is that these are functions whose properties are curiously analogous to those of the "circular" functions, but with the essential difference of being non-periodic. The student of physics will know that for this reason the hyperbolic functions are most useful in describing the behaviour of bodies whose movements would have been periodic but for the presence of some restrictive condition—for example, friction.

Stated, as above, in abstract terms the programme of the section may seem to threaten to be difficult and dry. If the argument were developed abstractly it might easily deserve both epithets. Fortunately it is possible to base the algebraic analysis of periodicity upon a concrete foundation of most attractive observations and experiments. Among these the most important are the movements of vibrating bodies, waves, and the fascinating phenomena of the tides.

The teacher may be reminded that the treatment of these topics in the mathematics lesson need be in no sense an intrusion into the field of physics; for the remark already made with reference to the study of astronomical topics in Section V applies also here. We are concerned only with those features of vibrations, waves, etc., which are visible to the eye of common sense, and our business is limited to the search for an adequate mode of describing what anyone may see who takes the trouble to look. The work of the physicist begins where our inquiry leaves off, for it is his business to bring to light the hidden conditions which determine the behaviour of vibrating and undulating bodies.

The reader whom these arguments leave unconvinced is invited to consider how greatly the whole subject of periodicity gains in rationality, as well as in interest, when it starts from a basis of organized observation. Consider, for example, the first step of the argument—the concept of an angle of endless magnitude. To the ordinary student there seems nothing more arbitrary and unnecessary than the extension of the angle-scale beyond 360°, and the resulting disturbance of the simple arrangement which associates a single angle with each single set of values of the ratios seems a purely gratuitous complication. But bring him face to face with the problem of finding a clear and compact mode of describing (say) the behaviour

of a swinging lamp and the aspect of the matter is entirely changed. He now sees that the extension of the angle-scale, instead of being a complication, is actually a simplification of the highest value.

This reason for the procedure adopted in Section VII seems so cogent that there is, perhaps, only one objection worthy of consideration. It may be protested that in spite of the foregoing disclaimer the study of vibrations and waves does as a matter of fact involve the introduction into the classroom of elaborate physical apparatus and experiments. The answer is that although when the teacher is a physicist as well as a mathematician he may with some advantage draw upon the resources of the laboratory, yet such a practice is by no means necessary, and, perhaps, on the whole, not even advisable. His object in the mathematical lessons is to instil mathematical ideas. It is claimed that these ideas cannot be effectively taught apart from the physical phenomena which they interpret, but it is part of the same argument that the phenomena should be studied in forms which are already familiar to everybody. Thus it will be found that where the teacher is advised to use a piece of apparatus it is always of the nature of a tridimensional diagram intended simply to help the pupil's analysis of the familiar phenomena under consideration. For this reason models are described which make the minimum demands upon the teacher's constructive ability. Moreover, since they are only diagrams they are not indispensable.

The subdivisions into which the exercises naturally fall have already been partly indicated. Exs. XCIX-CI introduce the notions of circular measure and of an angle as a quantity which may have any magnitude, positive or negative; they also investigate the application of the fundamental trigonometrical formulæ to angles of unlimited magnitude. Exs. CII-CIII generalize the preceding argument into a formal doctrine of the direct and inverse "circular functions". Exs. CIV-CVI apply the new notions to the study of wave-motion, including the composition and analysis of harmonic wave-forms and functions. Ex. CVII takes up the important question of the differential formulæ of the sine and cosine. Exs. CVIII-CIX treat of the hyperbolic functions and their analogies with the circular functions. Ex. CX is a supplementary exercise in which all these topics receive further

development, the most important topics being the calculation of π by series and the prediction of ocean tides.

§ 6. Section VIII. *The Theory of Limits.*—The special aims of this section have been foreshadowed in ch. xxiv., § 6, and are stated with some detail in the student's Introduction (*Exercises*, II, p. 321). The numerous arguments in which Wallis's Law and differential formulæ have been established or applied constitute, strictly speaking, only a "calculus of approximations". They are now to be placed upon a more satisfactory logical basis. Thus our first task in Section VIII is to convert the doctrine of approximations into the differential and integral calculus (in the proper sense of the terms) by means of the idea of a "limit". The second task is to develop a technique by which the methods of the calculus can be readily applied to the problems in which they are relevant. In view of former statements about the scope of this book it need hardly be said that the development is confined strictly to fundamentals, and that the range of the problems is restricted to those which may be considered of universal interest.

The general notion of a limit is explained in Ex. CXI and is applied in Exs. CXII-CXIV to the derivation of the standard formulæ of differentiation and integration. These exercises deal chiefly with the direct exemplification of first principles; secondary topics, such as partial differentiation, and practical applications, such as the theory of curvature, being reserved for the supplementary exercise. In Ex. CXV the use of the "differential formula," regarded as an approximation formula, is improved into the "differential equation," and some of the simpler applications of differential equations are illustrated. This part of the section is completed by a simple discussion of partial differentiation and integration in Ex. CXV.

Ex. CXVI is a connecting-link between the former and the latter subdivisions of the section. It raises the important theoretical question as to whether a function can always be differentiated and shows (following Weierstrass) that unassisted common sense is not a safe guide in the search for an answer. The important theorem called after Rolle finds its natural place here. In Ex. CXVII we consider the nature of expansions and ask whether it is possible to find a general rule for developing a function in powers of the variable—a

rule which shall include a means of estimating the degree of approximation to the true value of the function obtainable by a given number of terms. The answer to this demand is an investigation of the theorem generally known as "Taylor's". Ex. CXVIII supplements all the exercises of the section by examples of wider scope and greater difficulty.

§ 7. SECTION IX. *The Theory of Statistics.*—Examples and discussions which are, essentially, contributions to a theory of statistics have been met at various points of the course—Exs. XXVI, D, LXVI, LXVII; chs. IV., § 7, XXXVII., §§ 1, 2. The results of these are now to be gathered together and developed into a systematic doctrine embracing the fundamental ideas of the subject and illustrating some of its simpler and more important applications. Under the names "permutations and combinations" and "probability" certain parts of the doctrine of statistics have long had a place in the mathematical curriculum. It must be confessed that they hold it rather by hereditary privilege than by their own merits.[1] Another part—the theory of errors—is of great practical importance and theoretical beauty, but has not hitherto appeared in elementary text-books. In recent years, however, the science has undergone such remarkable developments and has been applied so widely and to matters of such immense importance that its claim to an honourable place in the school curriculum can hardly be resisted. A certain degree of cultivation of the "statistical sense" seems, in fact, likely to become one of the essential qualifications for intelligent citizenship. In any case it is no longer possible without it to understand modern developments in sciences, such as biology and psychology, which have hitherto been thought the refuge of the mathematically destitute. The aim of Section IX is to give in clear and simple outline an account of the general notions and fundamental technical methods of modern statistics. "Permutations and combinations" and "probability" have their proper places in this account, and will (like the periodic functions in Section VII) be found to gain very greatly in rationality and interest by being absorbed in a general doctrine of such far-reaching and substantial importance.

[1] Some important exceptions must be made to this general complaint. For example, Prof. Chrystal in his invaluable *Text-Book* gave to his treatment of probability just the turn here advocated.

The argument of the section shows three well-marked stages. The first (Exs. CXIX-CXXI) consists mainly in the revision and extension of the methods of recording the "frequency-distribution" of a series of measurements or other statistics. In the second stage (Exs. CXXII-CXXIII) we have a simple investigation of certain cases in which frequency-distributions can be calculated either from a knowledge of other frequency-distributions or upon the basis of certain assumptions about the nature of the things in question. At this point permutations and combinations are studied as examples of calculable frequency-distributions, and acquaintance is made with some of the theorems of the calculus of probabilities. In the last stage (Ex. CXXIV) we turn from the problem of calculating the frequency-distribution of a single variable in order to investigate the problem of "correlation," that is, the problem of estimating the mode and degree of dependence of one variable upon another. A supplementary exercise (CXXV) gives further illustrations of all the chief topics dealt with in the section.

§ 8. *The Minimum Course.*—Scheme B on p. 380 sets forth a programme of exercises for a minimum course to be taken by students who cannot work through the whole of the sections. The principle followed in selecting these exercises is to choose those which deal with subjects commonly required in public examinations, together with as much of the other work as is necessary to form a rational introduction to these subjects. In Section IV nothing is retained except the theory of indices, logarithms, and annuities, and the revision exercise. Section V is omitted altogether. In Section VI the two exercises are retained which contain enough material to provide a rational explanation of the "imaginary" roots of equations. Since a knowledge of these roots is (unfortunately) demanded at present even in elementary examinations, the two exercises in question are placed at the beginning of the scheme. In Section VII Exs. XCIX-CIII are all retained since they contain the theory of circular measure, of the trigonometrical ratios of angles of unlimited magnitude, the "sum and difference" theorems for such angles, and the theory of inverse circular functions—subjects which are found in the syllabuses of many school examinations. Nothing is taken from Section VIII, while Section IX contributes only Ex. CXXII and

Ex. CXXV, A, B, which deal with permutations, combinations, and their applications to the binomial theorem. The teacher who desires to do so will find no difficulty in filling out this minimum course either by more detailed treatment of the subjects of a single section or by adding exercises from several.

SECTION IV.

MAINLY REVISION.

THE EXERCISES OF SECTION IV.

*** The numbers in ordinary type refer to the pages of *Exercises in Algebra*, Part II; the numbers in heavy type to the pages of this book.

EXERCISE		PAGES
LXX.	INTEGERS	5, 405
LXXI.	RATIONAL NUMBERS	9, 409
LXXII.	IRRATIONAL NUMBERS	13, 410
LXXIII.	OPERATIONS UPON NUMBERS	19, 420
LXXIV.	THE COMPLETE NUMBER SCHEME	25, 421
LXXV.	FUNCTIONS OF ONE VARIABLE	32, 422
LXXVI.	SOME PECULIARITIES OF FUNCTIONS	40, 425
LXXVII.	FUNCTIONS OF TWO VARIABLES	47, 426
LXXVIII.	THE DEVELOPMENT OF ALGEBRAIC SYMBOLISM	55, 428
LXXIX.	ANNUITIES-CERTAIN	63, 432
LXXX.	CONTINGENT ANNUITIES	72, 432
LXXXI.	LIFE INSURANCE	76, 432
LXXXII.	THE EXPONENTIAL FUNCTION AND CURVE	79, 433
LXXXIII.	DIFFERENTIAL FORMULÆ	83, 435
LXXXIV.	SUPPLEMENTARY EXAMPLES	
	A-D. Revision papers 1-4	90, 438
	E. Scales of notation	96, 438
	F. Logarithmic approximation-series	97, 438

CHAPTER XXXIX.

NUMBER SYSTEMS.

§ 1. *The Scope of Exs. LXX-LXXIV.*—We have seen (ch. I., p. 5) that the object of an algebra is to develop a calculus, that is, a system of symbols and rules for the manipulation of the symbols, by means of which the investigation of some definite "province of thought or of external experience" may be facilitated. We have also seen that the nature of the symbolism and the rules of manipulation must in each case be determined by the special properties of the objects of thought or experience to which the algebra is to be adapted. In the case of ordinary algebra those objects are numbers. Thus everything in ordinary algebra is, in the first instance, a statement or a deduction about numbers or the operations which can be performed upon numbers. Since, however, numbers may, in one way or another, come into connexion with everything in the universe, an algebraic statement may refer, in the second instance, to any objects of thought or experience whatsoever.

But this statement requires amplification in one important particular, for it does not explain what is meant here by "numbers". The truth is that as we follow the development of algebra, either in actual history or in the teaching of the classroom, the meaning of the word "number" is itself found to develop. At the outset it means simply the ordinary integers and fractions of arithmetic supplemented, perhaps, by the "irrational" numbers that make their first, imperfectly understood, appearance as "surds". Section I of the present work was confined to the algebra of these signless or "non-directed" numbers. But the progress of the algebraic argument itself led to the introduction of a new class of numbers—positive and negative, or "directed" numbers. The investigation of the algebra based upon the

properties of these new numbers was the subject-matter of Sections II and III. At present the development has gone no further, but in Section VI it is to make us acquainted with yet another class of numerical entities—the "complex numbers" of the form $a + ib$. Ordinary algebraic theory has never found it necessary to go beyond these, and the exploration of their main properties marks the limit of the development of the idea of number in our course. It is, however, instructive to note that there is another algebra—the algebra of "vectors"—which is based upon yet another extension of the number-concept, and has, accordingly, laws differing in certain respects from those of ordinary algebra.

When we begin—as we propose to do in Section IV—to submit these numerical concepts to critical examination, the inquiry is seen at once to have a twofold character (cf. p. 16). On the one hand, the various types of numbers either entered historically into mathematical practice as means of furthering man's attempts to elucidate the behaviour of the external world or, having originated within the province of algebraic theory, were afterwards found to have useful practical applications. On the other hand, regarded simply as concepts or notions, they exhibit a definite logical development, fractions being derived from integers, irrationals from rationals, directed numbers from non-directed in accordance with ascertainable principles. Thus to understand numbers fully we must inquire both into their relations to the non-numerical entities with which they are connected in mathematical practice and into their logical relations to one another. Exs. LXX-LXXIV pursue, in a simple way, both branches of this inquiry, and are intended to present the most important results reached by Dedekind (*c.* 1872), Georg Cantor (*c.* 1883), and Bertrand Russell (*c.* 1903). The treatment is based, in the main, upon Mr. Russell's *Principles of Mathematics* (1903), supplemented at certain points by Prof. J. W. A. Young's *Fundamental Concepts of Algebra and Geometry* (1911). The teacher who wishes to follow the subject up should study those books, beginning with the latter. If he is at ease with a French book he should consult M. Louis Couturat's *Les Principes des Mathématiques* (1905), in which the results of the most important modern writers, including Mr. Bertrand Russell, are described with characteristic French lucidity. If he reads German he will do well to

consult Dedekind's famous pamphlet *Was sind und was sollen die Zahlen?* (1888), a work not of forbidding difficulty and one of the classics of the new logical movement. Some of Cantor's most important papers have been translated into French from the original German in *Acta Mathematica*, vol. ii. The translations include the famous *Grundlagen einer allgemeinen Mannichfaltigkeitslehre* (1883).

§ 2. Ex. LXX. *Integers.*—It is evident that the signless integers must have been the first numbers to be used by mankind, and also that they are the logical, as well as the historical, basis of all other systems of numbers. Our investigation naturally begins, therefore, with a study of their nature and properties. A given whole number, such as seven, can be looked at from two points of view. Regarded as a "cardinal" number its individuality consists in its connexion with all classes or collections of objects which have a certain specifiable property. The following is a possible but fanciful way of describing this property and so of discriminating from all other collections those with which the number seven is connected. On Sunday set aside or name or think of one member of the collection under examination and one only. Set aside or name or think of another on Monday, of another on Tuesday, and so on until Saturday. If by this time the whole collection is just exhausted it has the same cardinal number as the days of the week; if we call that number "seven" then its number is seven. It is evident that this definition by "one-to-one correspondence" gives to seven a standing which is quite independent logically of all other cardinal numbers. It would hold good if no one had ever discovered any other collections than those containing seven members. But, as a matter of fact, there are collections with other numbers, and these collections can be derived from one another by a definite process which may be endlessly repeated. To a collection with a certain number add another member and you obtain a collection with another definite number. When we take account in this way of the relations of seven to other numbers we are regarding it *ordinally*. In an elementary treatment it is not necessary to pursue this distinction very far. The important things to realize are (1) that integers are based logically upon the fact of one-to-one correspondence between collections, and (2) that they derive from the ordinal relations of different collections the property

which enables us to use their symbols as "labels" of the members of any sequence in which there is a definite order corresponding to the order which subsists among countable collections. In other words, we must realize that whereas integers, regarded as cardinals, are necessarily connected with things which have *magnitude*, regarded as ordinals they may be connected with any things which can be arranged in a sequence upon some definite principle, whether these things have magnitude or not. Thus the magnitude of a cardinal number is the magnitude of the collections which it describes, and is a property which it possesses, so to speak, in its own right. A number regarded ordinally may be said to be greater than another only because it implies a greater number of predecessors in the series of which it is a member. In view of the errors which follow from confusing these two totally different kinds of magnitude, it is best, except in special circumstances, to avoid speaking of ordinals as greater or less than one another, and to speak of them only as coming before or after one another in the sequence which they constitute. This remark applies especially to positive and negative numbers, for these are essentially ordinal (p. 160).

When the student has learnt that the notion of a cardinal number is really based upon the idea of one-to-one correspondence between the members of "similar" or "equivalent" collections, he is for the first time in a position to scrutinize profitably the notion of an "infinite" number. The beginning of the analysis is found in the recognition (§ 1, *b*) that an infinite number cannot be exhausted by counting. The best definition of an infinite collection is, however, supplied by the arguments of Nos. 4-8. The essence of all these cases is that part of a collection is found to correspond, term by term, to the whole collection of which it is a part. It follows that an infinite collection cannot be obtained by adding term to term—nor be destroyed by taking away term after term; but this property is best regarded as a consequence of the former, which is to be taken as the definition and test of an infinite collection.

The objections which common sense suggests to this startling notion of an infinite number are in part answered by the last paragraph of the exercise. The natural man starts out with the prejudice that all numbers must be like the finite numbers whose properties can be explored by

counting. Thus he finds it paradoxical to admit the existence of numbers some of whose properties are radically different from those with which he is familiar. Faced with incontrovertible arguments, like those of Nos. 4-8, he is apt to turn philosopher and regard them as demonstrating nothing but the inability of the human intellect to reach absolute truth! There are two ways of dealing with this pessimistic conclusion. The first is to show that the rejection of infinite numbers in the sense defined leads common sense itself to ridiculous conclusions. An example of this form of reply is given in Ex. LXXII, No. 4. The second is to show, as the modern mathematicians have shown abundantly, that the assumption that in some collections the part is actually equivalent to the whole leads to no results inconsistent with any other known truths. That is to say, unbelievers must be dealt with much as a mathematician who lived in a world where only rectilinear figures were to be seen might deal with persons who ridiculed his definition of a new figure which he called a circle.

The method of one-to-one correspondence can be used to show that there are different infinite, just as there are different finite, numbers enjoying each its distinct individuality. Thus it is possible to imagine any number of sequences whose members have one-to-one correspondence with all the integers, just as all collections whose number is seven have one-to-one correspondence with the days of the week. The days of a man who is born but never dies are an example of such a sequence. If we may suppose that it is literally true that "of the making of books there is no end," and that no two books are ever finished at exactly the same time, we have another example in the succession of books, beginning with the first inscribed clay tablet but never ending. The characteristics of all such sequences are (i) that there is a definite first member of the sequence, (ii) that there is no last member, and (iii) that every member has a definite successor. It is because they all possess these characteristics that the various sequences of this kind exhibit one-to-one correspondence between their terms. It follows that they must all be supposed to have the same infinite number. This number cannot, of course, be expressed in digits but it may (like the finite numbers e and π) be referred to by a symbol. Cantor used the Hebrew letter Aleph with zero as suffix to indicate that it is

the first of the infinite numbers. Other writers prefer the symbol a_o. Next, it is easy to point to other infinite collections whose terms cannot be put into one-to-one correspondence with the natural numbers. The points on a straight line of finite length offer an example. They are infinite in number, for a part of the line can be put into point-to-point correspondence with the whole (Ex. LXXII, No. 3). On the other hand, they differ from the sequence of integers in two respects: (i) there is a last member as well as a first; (ii) no member can be said to have a definite successor, for between any two points, however close, there is always a third. On account of these differences the points cannot be put into one-to-one correspondence with the integers. Nevertheless, Cantor has shown [1] that there are at least two other collections with which they *can* have one-to-one correspondence—namely the points of a square and the points of a cube. Thus, just as the collection of all the integers may be regarded as the standard example of the infinite number a_o, so the collection of points on a straight line may be taken as the standard instance of another infinite number—generally called the "number of the continuum". It can even be shown that the two are connected by the relation

$$\text{number of the continuum} = 2^{a_o}$$

The analysis of the idea of infinite number leads to a practical conclusion which the teacher should constantly apply. The value of a variable is often said to be "infinite" when it is greater than any finite number that can be named, that is when it has no maximum. This is an improper use of the term, for it does not follow from the fact that it has no maximum that it is the number of a collection of which a part is equivalent to the whole. All that can be said is that it is indefinitely or endlessly great, and the symbol "∞" should be understood to mean this and should never be read as "infinite". It is still more desirable to avoid the too common statement that the quotient of a finite number by zero is infinite. As we have seen at earlier points of the book (pp. 123,

[1] Young, *Concepts*, etc., Lectures VIII and XVI; Russell, *Principles*, p. 311; Cantor, *Acta Math.*, II., "Une contribution à la théorie des ensembles". See also Ex. CXVI, No. 20, and ch. LII., § 1.

376) this statement is simply nonsense. In fact the discussion of this article will have served a useful purpose if it convinces any reader for the first time not only that the word "infinite" is a numerical term capable of exact definition and therefore of precisely limited application, but also that the occasions for applying it in elementary mathematics are comparatively rare. It is too apt to be used at other times merely as a cloak to conceal absence of precise thought.

§ 3. Ex. LXXI. *Rational Numbers.*—All the other numbers which appear in mathematics are based upon the primitive series of integers. The first of these "artificial" numbers are "fractions". A fraction is, strictly speaking, a *pair* of integers associated in accordance with a definite law. This law enables us to substitute for each single integer a pair of integers which can be taken as equivalent to it—the pair being the given integer itself and the integer 1. (For example 3/1 is equivalent to 3.) In this way we obtain an infinite collection of numerical entities all of the same form. They constitute the so-called "rational numbers".

Nos. 1 and 2 are meant to bring out the differences between the sequence of integers and the sequence of rationals arranged in their "natural" order. In **No. 1** we see that there is an infinite number of rationals between any two members of the series. This fact is evident from the consideration that the sequence is "dense" (p. 10) so that its terms cannot be exhausted by counting. In **No. 2** (ii) the number of rationals between 10 and 20 inclusive is infinite, there is a first term (10) and a last term (20). It cannot be said, however, that any term has either an immediate predecessor or an immediate successor; for example, it is impossible to name any definite fraction as coming next in magnitude either before or after the fraction $\frac{3}{8}$. In **No. 2** (iii) the rationals *between* 10 and 20 have no first term and no last term; for the rational 10/1 has no immediate successor and the rational 20/1 no immediate predecessor. In **No. 3** every term has an immediate successor, so the sequence cannot be dense. On the other hand, not every term has an immediate predecessor; for it is clear that (since there is no limit to the value of r) the groups of terms preceding the terms 2, 3, 4, . . . have no last member.

When we take the symbols of the rationals and, breaking away from the "natural" order, arrange them in the order

indicated in **Nos. 4, 5**,[1] we obtain a sequence which is "ordinally similar" to the sequence of integers. That is to say, it has a first but no last term, every term except the first has an immediate predecessor and every term an immediate successor. It follows **(No. 6)** that the rationals arranged in this way can be brought into one-to-one correspondence with the integers in their natural order, and therefore have the same infinite number, a_0. To use Cantor's term, they form a denumerable series. This result is a striking instance of the power of the conception of one-to-one correspondence.

§ 4. Ex. LXXII. *Irrationals. Continuity.*—The subject of this exercise came before us at an early point of the course. It was seen in ch. VIII., B (p. 94), that, although a number may always be obtained whose square is as nearly as we please equal to a given number, yet in most cases there seems no probability that an exact square root could be found. Consequently, when we came in ch. XXIII. to identify the complete sequence of positive and negative numbers with points on an endless straight line, we were obliged to recognize that certain points correspond to no number at all, integral or fractional. For example, the point whose distance from the origin measures the length of the side of a square whose area is 2 units would be such a point. If this point is to have a numerical label at all, we must invent a new one. It was suggested (p. 231) that the best label for this purpose would be the symbol "$\sqrt{2}$".

Now the interesting thing about this suggestion is that the symbol $\sqrt{2}$ had hitherto been used to represent any number whose square was sufficiently near to 2 for the purpose in view (p. 94). Thus, regarding the line OX in *Exercises* II, fig. 62, as the positive half of the linear scale referred to above, $\sqrt{2}$ would not be the label of the point P, but could be attached ambiguously to any point sufficiently near to P. It follows that the proposal to use $\sqrt{2}$ as a label for P itself is, strictly speaking, a new departure. When examined more carefully the proposal is seen to raise the following important question: Symbols such as 7 or $\frac{3}{8}$ are not only labels for specific points on the line OX but are also symbols for *numbers;* can it be said that $\sqrt{2}$, which we have now

[1] The scheme in No. 5 is copied from Young, *Concepts*, etc., p. 74.

adopted as the label for the point P, is also the symbol for a number? The problem of Ex. LXXII is to answer this question and the subsidiary questions which it includes.

The solution of the problem is very modern, but the problem in its essence is of great antiquity. Pythagoras (*c.* 530 B.C.) or his followers discovered that the diagonal of a square is incommensurable with the side—that is, that the ratio of the lengths of the lines cannot be expressed by any integral or fractional number. This discovery was the starting-point of investigations that culminated in Euclid's Book X, a masterly treatise upon "irrational" ratios.[1] In accordance with a remark made elsewhere (pp. 32, 70), Euclid's argument, though essentially arithmetical, was necessarily conducted in geometrical terms. A treatment arithmetical in form as well as in substance was possible only after the invention of the Arabic notation.

One of the earliest examples of such a treatment is to be found in the second book of the *Arithmetica Integra* (1544) of Michael Stifel, "pastor of the Church at Holtzdort".[2] It is practically a restatement of Euclid's doctrine in the Arabic notation, together with a running commentary upon it. In the first chapter he comes at once to the question raised above: namely, whether the so-called "irrational numbers" are really numbers at all. As he subsequently points out (ch. II.) Euclid was clearly of opinion that they are not numbers.[3] In this chapter, however, he discusses the matter, as we have done, with reference to the problem of expressing "irrationals" in the decimal notation. Some of his observations are worth quoting as exemplifying the difficulties and uncertainties that beset all early attempts to give a clear account of the fundamental notions of arithmetic and algebra (cf. p. 159). On the

[1] See Sir T. L. Heath's edition, Vol. III. It is interesting to notethat Euclid's use of the terms "rational" and "irrational" is not identical with the modern one. According to his definition a line is irrational with regard to a given unit-line only if neither the lengths of the lines nor the areas of the squares upon them are commensurable.

[2] The table of binomial coefficients given on p. 213 appears in Book I of this work (folio 45).

[3] Book X, Prop. V: "Commensurable magnitudes have to one another the ratio which a number has to a number"; Prop. VII: "Incommensurable magnitudes have not to one another the ratio which a number has to a number".

one hand, argues Stifel, "since, in proving geometrical figures, when rational numbers fail us irrational numbers take their place and prove exactly those things which rational numbers could not prove, . . . we are moved and compelled to assert that they truly are numbers, compelled, that is, by the results which follow from their use—results which we perceive to be real, certain, and constant". "On the other hand, other considerations . . . compel us to deny that irrational numbers are numbers at all. To wit, when we seek to subject them to numeration . . . we find that they flee away perpetually, so that not one of them can be apprehended precisely in itself. . . . Now that cannot be called a true number which is of such a nature that it lacks precision. . . . Therefore, just as an infinite number is not a number, so an irrational number is not a true number, but lies hidden in a kind of cloud of infinity." "Again," he continues, "if irrational numbers were real numbers they would be either whole numbers or fractions." They are certainly not whole numbers, for it is easily seen that each irrational falls between

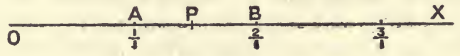


FIG. 92.

two consecutive integers. Also they are not fractions with definite numerators and denominators, for when a fraction is multiplied by itself it can never yield a whole number, while an irrational, if it is the square root of an integer, will always do so. Thus they cannot be real numbers.

It is not easy to improve upon this statement of the dilemma. On the one hand, irrationals must be real numbers; for calculations which make use of them are just as trustworthy as those which use only rationals. On the other hand, they cannot be real numbers because they are neither integers nor fractions. There is only one way of escape from it; some definition of the term "number" must be found which, while making rationals a special case of a general concept, will at the same time leave a place for irrationals as another special case.

The preliminary discussion in ch. XXIII. suggests two lines of attack of which one or both may lead to a conquest of this problem. Let OX (fig. 92) be a linear scale, beginning with the point O but endless towards X. Imagine every

point which corresponds to a rational number to be labelled with the symbol of that number, integers being represented, for the sake of homogeneity, in the form of fractions with 1 as the denominator. Since it is impossible actually to show all these numerical labels in the diagram, we content ourselves with inserting three as specimens ($\frac{1}{1}$, $\frac{2}{1}$, $\frac{3}{1}$). Let P be any point on the scale selected at random; then, by hypothesis, P is in every case to be regarded as the representative of a number. If we happen to have hit a point to which one of our labels is attached, there is, of course, no difficulty; the number represented by P is simply the number described by the label. Trouble arises only if the point selected has no label; for then we have to ask how this point can be regarded as representing a number *in the same sense as the point which does bear a label.* To answer this question we note that the point P can be thought of in two ways: first, it marks a definite position on the line OX between points on the left of it and points on the right; secondly, it marks a definite length of the line OX, starting from the origin. Both of these statements are true of P whether it is or is not labelled with the symbol of a rational number; either of them may be taken, therefore, as the principle underlying the new definition of the word "number" which we are seeking. Since the new meaning of "number" is to include more than "rational number" includes, it will be well to associate with the word a distinguishing adjective, just as we added the adjective "rational" when we wanted "number" to mean more than the original sequence of integers. For historical reasons the term "real number" must be adopted for this purpose in spite of its misleading suggestions (ch. XLIX.). Our immediate task, then, is to find and to examine two alternative definitions of "real numbers," both having the property that they include two subclasses of numbers, "rationals" and "irrationals," upon an equal footing.

In seeking the first definition we follow the path taken independently by the German, Richard Dedekind, and the Frenchman, Paul Tannery. Let us begin by supposing the point P (fig. 92) to be associated with a rational number—for example $\frac{3}{2}$. Looking at this number isolated from others we may regard it, as in § 3, simply as a combination of the integers 3 and 2. On the other hand, if we look at the symbol $\frac{3}{2}$ in

its place among the other symbols disposed along the line OX, we see that, just as P may be regarded as a boundary separating points to the right of it from points to the left, so the number symbolized by $\frac{3}{2}$ may be thought of as a boundary separating the numbers before it in the natural sequence from the numbers that follow it. It may help to keep these two aspects of the number apart if we adopt different ways of printing its symbol to correspond to them respectively. Thus we may print the symbol in ordinary type, $\frac{3}{2}$, when we think of it simply as a combination of integers, and in heavy type, $\boldsymbol{\frac{3}{2}}$, when we think of it as a boundary between the pairs of integers below and the pairs of integers above. In the second capacity $\boldsymbol{\frac{3}{2}}$ is to be called a "real number".

Fig. 93 illustrates this definition. The point P is shown isolated, the two segments, OA and BX, which it separates, being withdrawn to right and left, together, of course, with the number-symbols attached to them. It is important to observe that the segment OA can have no definite end-point to the right, for in the original line it was impossible to say

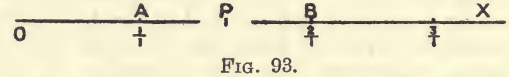


FIG. 93.

that P had a definite point immediately to the left of it. Similarly, the segment BX has no first point at the end adjacent to P. Thus the point P may be regarded as a "cut" (*Schnitt, coupure*) which divides the whole line into a lower segment which has no end-point and an upper segment which has no point of beginning.

An alternative way of expressing the same facts is to say that P is the "upper limit" of the segment OA and the "lower limit" of the segment BX, P being itself excluded from both these segments. Correspondingly, we may say that the "real number" $\boldsymbol{\frac{3}{2}}$ which is associated with P is a "cut" in the complete sequence of rationals taken in their natural order, a cut which divides them into a lower segment without a last member and an upper segment without a first member. Alternatively, we may say that it is the upper limit of the rationals (i.e. of the pairs of integers) which come before $\frac{3}{2}$ and the lower limit of the rationals which follow $\frac{3}{2}$.

Now suppose that our random choice had lighted upon a point P which is not associated with a label. Then we cannot

this time say that the point corresponds to a number in the sense that $\frac{3}{2}$ is a number. But we may nevertheless think of it as corresponding to a number in the sense that $\frac{3}{2}$ is a number; for it marks in exactly the same way as before a "cut" dividing the complete sequence of rationals into an endless lower segment and a beginningless upper segment. Since we have decided that the term "real number" shall imply simply this aspect of being a "cut" or a "limit" in the sequence of rationals, we are entitled to say that although P does not now correspond to a rational number yet it still corresponds to a real number. It remains only to assign to this number (i.e. to this mode of section of the complete sequence of rationals) a suitable name and a suitable symbol.[1]

To sum up: Imagine fig. 93 completed by the addition of the symbols of all the rational numbers each in its proper place in the scale. Then the figure represents a division of these numbers into a lower segment without a last term, and an upper segment without a first term. Each of the infinitely numerous ways in which this division can be made constitutes a "real number". It may be that the point P itself corresponds to a rational number which is therefore excluded from the two segments. In this case the mode of division is a "rational real number" and is named from the excluded rational represented by P. Again it may be that *all* the rationals fall into either OA or BX. In that case the mode of section is an "irrational real number," and must be named in some suitable way. For example, if the rationals in the lower section are all those whose squares are less than two, and the rationals in the upper section all those whose squares are greater than two, then the "real number" is most conveniently called "the square root of two". Lastly, a rational number may be thought of either as merely a rational number or as a rational real number; an irrational must always be thought of as a real number.[2]

[1] Compare with this the argument on p. 182 which established the validity of such expressions as $4-7$, in which the second number is greater than the first.

[2] The terms "rational real number" and "irrational real number" are used, for convenience only, to distinguish real numbers which correspond to rational numbers from those which do not. It is obvious that there is no difference in "rationality" between the two subclasses.

We turn now to the definition of a real number suggested by the consideration that the number attached to the point P may be thought of as representing the length of the segment OP. Here we follow Mr. Bertrand Russell. Once more divide the linear scale into two segments, but let P be the last point of the lower segment (fig. 94). All the symbols of the rationals will now always appear in one segment or the other. There will, however, still be two cases. In the first case the point P which ends the lower segment corresponds to a rational number in the sense denoted by the symbol $\frac{3}{2}$. In this case the corresponding "real number" represented by $\frac{3}{2}$ may be defined as the whole collection of rationals whose symbols would have their places to the left of P. In the second case the point P will represent no rational. But in this case also it will be considered as corresponding to a real number, and that real number will again be defined as the collection of all the rationals whose places lie to the left of P. Thus, according to Mr. Russell's definition, every real number is a collection or set of rationals taken in order from zero up-

FIG. 94.

wards. In some cases the set can be described adequately by the statement that it consists of all the rationals which are less than a certain rational N. In that case the real number is a rational real number and may be named from the rational N, though it would be well to emphasize the difference between N and the real number named from it by a distinctive method of printing the symbol (**N**). In other cases there is no definite rational N from which the set may be named. In that case the real number (i.e. the set of rationals) is an irrational real number and must be named in some suitable way. For example, if the set consists of all rationals whose squares are less than two it may conveniently be called the square root of two.

The Dedekind-Tannery definition has been explained fully, partly because of its historical importance, partly because it is at present the one most widely known, partly because it involves ideas—such as a "limit"—which will be of importance in the sequel. There can be no question, however, that for a student who approaches the subject for the first time

the Russell definition is incomparably the easier to grasp. Moreover, it is logically superior—and this fact is without doubt the source of its greater simplicity. For it will be seen that rational and irrational numbers as defined by the former method are not really homogeneous. When the limit of the two segments is a rational number it is actually there; when it is not a rational number it is not there, but has, so to speak, to be defined into existence. On the other hand, the sets of rationals which constitute real numbers according to the Russell definition are equally "there" whether they correspond to rationals or not.

For these reasons the Russell definition is taken as the basis of the treatment in Ex. LXXII. It need hardly be said that the illustration of the boxes is no part of Mr. Russell's argument, but is introduced merely for didactic purposes and, in particular, to show how a "continuum" of numbers can be built up from the ordered sequence of integers (through the intermediate step of a "compact" sequence of rationals) without reference to a line or any other magnitude. If the teacher does not find the illustration helpful, he can, of course, discard it and substitute a treatment more on the lines of the present article. With regard to the idea of continuity itself it would probably not be profitable or even possible to carry the discussion much farther than it is carried in Ex. LXXII. The results reached there are (i) that the points of a line are a sequence obeying Dedekind's Postulate, (ii) that the same is true of the real numbers, and (iii) that for this reason both are said to possess continuity. It should be noted that we do *not* prove that our scheme provides a real number for every point on the line. Whether it does so or not it is apparently impossible to say. What can be said is, however, that mathematicians have never found it necessary to postulate any points to which real numbers do not correspond. It is worth while examining a little the significance of this historical fact. Take one of the imagined boxes of Ex. LXXII, § 2, and suppose it labelled with an irrational symbol I. This label implies that the box is reserved for a perfectly definite collection of rationals (constituting the real number I) and that all other rationals are to be placed elsewhere. Now suppose the label to be changed to another irrational symbol I′ whose place is higher up the linear scale. Then it is clear that the box now offers hos-

pitality to rationals which were previously excluded; for I′ would not be a different real number from I if it consisted in the same set of rationals. Thus we must suppose that certain rationals which were previously outside are now to be found inside the box. But it is clear that any one of these rationals (say R) might have stopped on its way from the outside to the inside and itself become the label of the box. The set of rationals implied by this new label would, of course, be the real number R. It appears, therefore, that between any two real numbers I and I′ which are not associated with rationals there must be at least one real number which is so associated. Now the same thing may be true of the points on a line; in that case real numbers, defined as collections of rationals, would suffice for any calculation concerning points on the line. On the other hand, the constitution of a line might be such that two points unconnected with rationals could be found with no point connected with a rational between them. In that case real numbers would not be adequate for all calculations about lines. But as a matter of fact no argument has ever been produced which contradicts the common assumption that the real numbers form a number scheme adequate for all calculations involving space.

The teacher should not fail to point out that the results of Ex. LXXII are needed to put much of the work of Part I upon a proper logical basis. Thus it has constantly been supposed that the graphs of functions such as $y = ax$, $y = a\sqrt{x}$, $y = a^x$ are continuous lines, i.e. lines unbroken by any gaps in which points could be inserted which were not points of the original graph. Yet if the graphs are only assemblages of the points which correspond to rational values of x and y there must be an infinite number of such gaps. The fact that they cannot be exhibited to the eye does not touch the fact that reason shows them to be there. Thus all through our elementary work we have been tacitly assuming Dedekind's Postulate with regard to the number-scale; that is, we have taken it for granted that there is always a number corresponding to any point upon a line. We now know that this assumption is justifiable only if by number we mean "real number".

§ 5. *The Nature of* e *and* π.—The special position of the numbers e and π demands some mention in a review of numbers in general, but it must be restricted here to a brief

note. The question whether it is possible to "square the circle" has, of course, had a very long history; that of the value of e goes back only to the seventeenth century. In the case of both these entities the inquiries of mathematicians have gradually narrowed down to the question whether they are "algebraic" or "transcendent" numbers. A number is algebraic if it can be the root of an equation of any degree

$$C_0 + C_1x + C_2x^2 + C_3x^3 + \ . \ . \ . \ + C_nx^n = 0$$

in which the coefficients C_0, C_1, C_2, etc., are rational numbers. If it cannot be such a root it is transcendent. The solution of the problem is quite modern. In 1873 Hermite succeeded in proving that e is transcendent. In 1882 Lindemann followed with a proof of the transcendence of π based upon the results of Hermite. The reader who wishes to know the nature of these proofs should consult a paper by Prof. D. E. Smith in Young's *Monographs on Modern Mathematics*. The elementary student must be content to know that although e and π are not rational numbers their "irrationality" differs in an important respect from that of surds; these are algebraic numbers, those are not.

§ 6. *The Paradoxes of Zeno.*—We now turn to the examples of Ex. LXXII. In **Nos. 2, 3** we simply meet again the property that in an infinite collection a part can have one-to-one correspondence with the whole. **No. 4** states the best known of the paradoxes of Zeno. These paradoxes have generally been thought to prove the incapacity of the human mind to deal with the idea of the infinite, but Mr. Russell has shown that they cease to be insoluble riddles when we recognize that infinite numbers have the property to which reference has just been made. The argument underlying the paradox is as follows: At any moment of time Achilles and the tortoise are each at some point of their respective paths. Thus during any given period (that is, in the course of a definite series of moments) each must visit the same number of points. Hence the path of the tortoise cannot be a part of the path of Achilles, for it would in that case contain fewer points. The fallacy lies, of course, in the last statement; for although one path is a part of the other the number of points in each is the same—namely, the number of the "real numbers". Thus, even though Achilles

should far outstrip his competitor, the points occupied by each in his course can be correlated one by one.[1]

§ 7. Ex. LXXIII. *Operations upon Numbers.*—The aim of this exercise is to find definitions of the arithmetical operations which will apply to numbers of all kinds. The analysis is probably carried far enough there for the average pupil but may be completed here.

Addition of integers is explained in the exercise by the combination of sub-collections into a whole; addition of rationals by the combination of lengths. It will, however, be seen that if the symbols are taken to represent real numbers then the things combined are once more collections —to wit, collections of rationals. The same would be true if the symbols were irrational. Thus if we take "number" to mean "real number" there is a single definition for addition and (therefore) for subtraction.

Similarly for multiplication and with it division. It is easily shown that "repeated addition" is a definition which holds good only for integers, but that the alternative definition is applicable to all numbers if they are regarded as sets of rationals—that is, it is applicable to all real numbers. The argument of Ex. LXXIII, § 4, is directed to the case of rationals, but it is obvious that the definition of the product would apply equally if the factors were irrational.

In the case of the sum and product of rationals and irrationals there is a second problem. It is not enough to find a formal definition which brings them into line with the sums and products of integers; we must also find *numbers* to express them. The case of rationals is dealt with in §§ 3, 4. The principle followed is that the sum and product are to be measured by rules that would give, in the case of rationals which are also integers (e.g. 6/1, 17/1), the numbers which we already recognize as the sum and product of those integers. In the case of the sum of two irrationals the rule gives us no assistance; for example, if the irrationals are $\sqrt{a}$ and $\sqrt{b}$, the sum can only be expressed in the form $\sqrt{a} + \sqrt{b}$. In the case of products it may, however, often be applied with advantage. Thus let $\sqrt{a}$ and $\sqrt{b}$ be any two real numbers, rational or irrational, and let the rectangle

[1] For this and the other paradoxes see Russell, *Principles*, ch. XLII.

be drawn the points of whose area represent the pairs of rationals taken from the sets denoted by $\sqrt{a}$ and $\sqrt{b}$ respectively. Then the top right hand point will be $(\sqrt{a}, \sqrt{b})$, and our problem is to replace this pair of numbers by a single number. Now in every case in which a and b are squares the area will be measured by $\sqrt{c}$, where c is the product of ab by the rules for integers or rationals. By our principle, therefore, $\sqrt{c}$ is also to be counted as the product when $\sqrt{a}$ and $\sqrt{b}$ are irrational. This argument solves **No. 14.** The teacher should have no difficulty in finding another instructive proof based on the idea that the product of $\sqrt{a}$ and $\sqrt{b}$ must be represented by a number which comes between (i.e. is the "limit" of) the products pq and $p'q'$ where p and q are any rationals less than $\sqrt{a}$ and $\sqrt{b}$ respectively, and p' and q' any rationals greater respectively than these irrationals.

§ 8. Ex. LXXIV. *The Complete Number Scale.*—A good deal of the argument of this exercise has already appeared in ch. XVII., §§ 1-4, and in chs. XVIII. and XXIII. To these the reader should refer. Attention should be given to § 2 of the exercise, for it expresses the essence of algebra regarded as a symbolic logic: Algebraic operations simply carry us backwards and forwards along the endless linear continuum of the numerical symbols, but we can always interpret our results in terms of things (numbers, magnitudes, etc.) which have independent existence. In Section VI this view will be modified to the extent that we shall find that the operations of algebra carry us over a continuum of symbols of two dimensions—namely, the field of "complex numbers"—but in essence the present conclusion will stand.

In this exercise the student is introduced to the interesting and important idea of a "group". The theory of groups is a large department of pure mathematics which has very important applications in mathematical physics. The account given in the exercise is based on Young, *Concepts*, etc., Lecture IX.[1]

[1] Prof. E. W. Hobson's little book, "*Squaring the Circle*" (Camb. Univ. Press), appeared too late to be mentioned in § 5. The teacher will find it an admirable, as well as an authoritative, introduction to its subject.

CHAPTER XL.

FUNCTIONS.

§ 1. Ex. LXXV. *Functions of one Variable.*—The first two articles of this exercise are mainly a restatement in print of what the student has learnt orally in Part I; the third gives a simple explanation, in graphic terms, of the meaning of the adjectives "continuous" and "discontinuous" as applied to functions. The teacher should remind the class

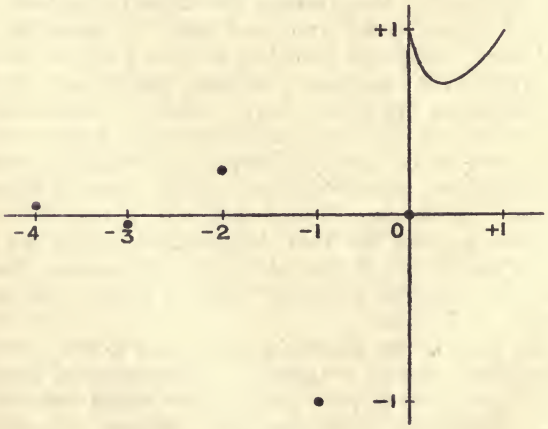


Fig. 95.

that in speaking of y as a continuous function of x we presuppose that the values of x themselves form a "continuum"; i.e. that x means any and every real number within the range in question. If this were not the case the graph could not be considered unbroken even though the function were a continuous one (p. 418).

Fig. 95 is the graph of **No. 12.** The dots on the left of

the y-axis correspond to $(-4)^{-4}, (-3)^{-3}, (-2)^{-2}, (-1)^{-1}$. These are isolated values of the function, for it is impossible to find a real number which gives the value of x^x when x is both negative and non-integral. After $x = 0$ the function is continuous, + 1 being the "limit" as x approaches 0 from the positive side. Figs. 96-8 are the graphs of the functions

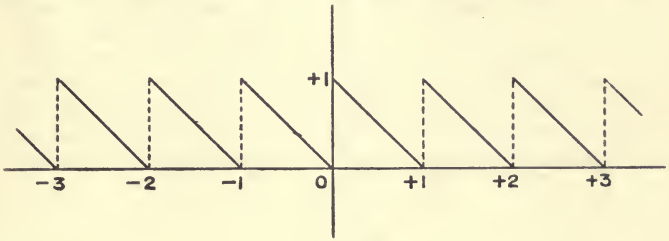


FIG. 96.

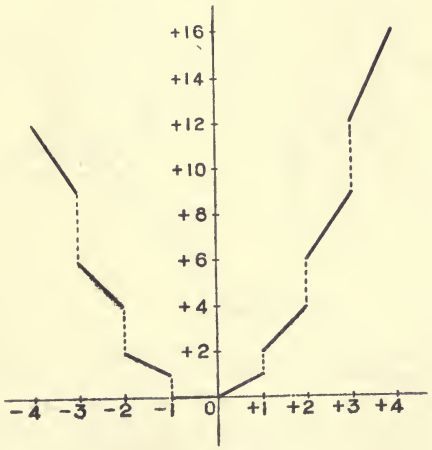


FIG. 97.

of **No. 13** (cf. Hardy, *Pure Mathematics*, p. 47). Fig. 115, p. 484 (omitting the broken line), is the graph of **No. 14.**

In division B the examples deal with the problem of determining functions which have given properties. Incidentally the student learns the use of the important notation

$y = f(x)$. The interesting examples **Nos. 21-3** were suggested to the author by Dr. L. Silberstein; fig. 99 is the solution of **No. 23**. It is obvious that the graph could be continued in "waves" increasing indefinitely towards the

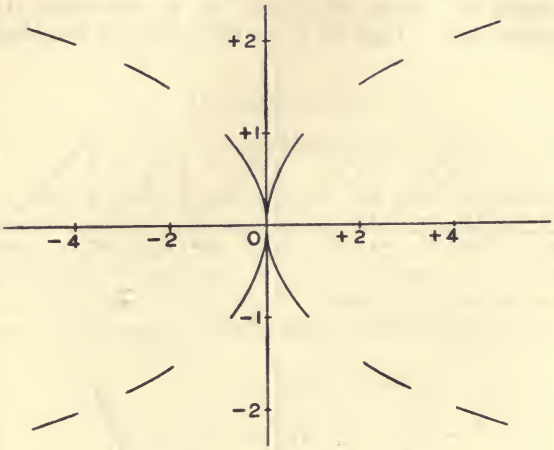


FIG. 98.

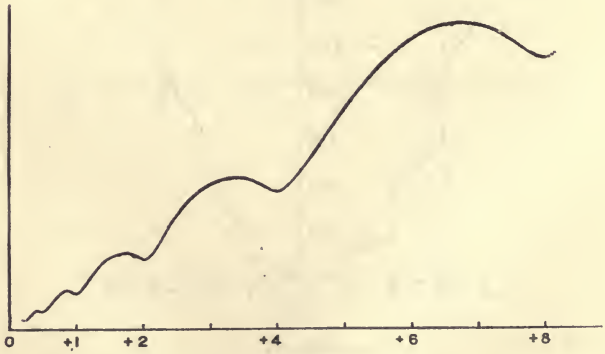


FIG. 99.

right and decreasing indefinitely towards the left. The same remark applies to the graph of **No. 28**. Note the parallelism between Nos. 21-3 and Nos. 27-8. **Nos. 31-2** are taken from Mr. G. H. Hardy's *Pure Mathematics*. In **No. 31** the graphs approximate to fig. 100 A, in **No. 32** to fig. 100 B.

§ 2. Ex. LXXVI. *Some Peculiarities of Functions.*—The first division of the exercise is again for revision. Careful attention should be given to the argument in **Nos. 8-11.** In **No. 11**, putting $x = 1 + h$, we have

$$y = m + \frac{m(m-1)}{2!}h + \frac{m(m-1)(m-2)}{3!}h^2 + \ldots$$

whence it is obvious that as h approaches zero (i.e. as x approaches unity) y approaches m.

Division B revises the notion of a "gradient" and generalizes it into the idea of the rate of change of a function as the value of the independent variable changes. The teacher may find it profitable to treat this topic by means of Newton's conception of "fluxions". Newton thought of

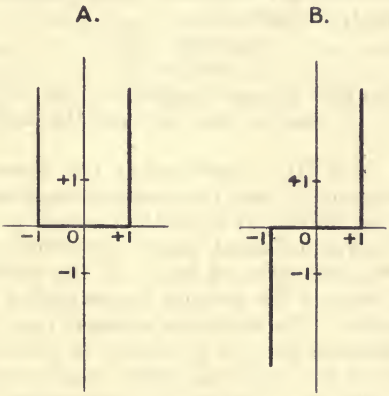


Fig. 100.

both variables as changing with the time; the rate of change of the function is, upon this view, the ratio of the velocities with which the two variables are increasing. It is possible that this way of looking at the matter was suggested to Newton by Napier's theory of logarithms. We have seen (p. 300) that Napier thought of the sine and its logarithm as represented by points (fig. 77) moving along two parallel lines. This idea may be generalized and the points regarded as representing the values of x and y in the case of any function $y = f(x)$. Then the speeds of the two points are the "fluxions" of the variables, represented by Newton by

the notation $\dot{x}$ and $\dot{y}$. The rate of change of the function would thus be the ratio $\dot{y}/\dot{x}$. If the lines along which the points are moving are set at right angles to one another the successive positions of the points determine the graph of the function as we understand it—just as the growth-curve of fig. 78 (p. 302) is generated from fig. 77.

In dealing with this subject the teacher should bear in mind what was said on pp. 246-50 about the relation of the "calculus of approximations" to the doctrine, based upon the theory of limits, to be studied in Section VIII.

Division C offers a simple treatment of the fascinating topic of "singular values". In **No. 25** the graph must first be shifted a unit place to the left so that the point (+ 1, 0) may become coincident with the origin. The corresponding expression for the function is

$$y^2 = (x + 1)x^2$$
$$= x^3 + x^2.$$

It is now seen that the two tangents at the origin are given by $y^2 - x^2 = 0$; that is, that they are the lines $y = x$ and $y = -x$.

§ 3. Ex. LXXVII. *Functions of two Variables.*—From the technical point of view this exercise has great importance as teaching the student to apply the methods of "coordinate geometry" to tridimensional space. In accordance, however, with the principle explained on p. 47 the argument is presented as a study of the graphic representation of functions of two variables. The treatment is based upon an idea with which the student may be presumed to be fairly familiar, namely, that of the "contour lines" which are used to represent the relief of a district in a geographical map. The teacher should note (i) that the general notion of a function of two variables is approached through the study of concrete cases (cf. pp. 109-10 and ch. XII.); and (ii) that constant use is made of the two devices of shifting and rotating a graph in order to obtain the functions to which it corresponds in its various positions (see ch. XXVI., A, § 2; B, § 2; and *Exercises*, I, p. 325).

The student should be taught to represent and to realize solid forms by means of contour lines just as in the case of the map of a mountainous country. It should be understood that in order to obtain an adequate idea of a given form it is sometimes necessary and always helpful to draw "contour

maps" of it corresponding to sections parallel to each of the three coordinate planes. These notions are suggested in **Nos. 9-13.** It is also well worth while to construct at least one solid from its contours in the way indicated in **No. 13.** A simple method is to cut a series of contours out in paper or thin card and to use them as "templates" in cutting slabs of the same shapes out of a sheet of thick but tractable material. The felt sold to fix under carpets serves well for this purpose; it is cheap, uniformly thick, and easily manipulated. When the slabs have been fixed in the positions in which they build up the solid, the surface may be smoothed by means of a sharp knife or razor. Clay or plasticine models are, of course, better than those of felt but are more troublesome to construct.

The student who has worked Part I, Exs. LXIV and LXV, is not likely to find difficulty in the examples of the present exercise. If Exs. LXIV and LXV have not been worked they should be taken between Exs. LXXVI and LXXVII.

CHAPTER XLI.

THE EXPONENTIAL FUNCTION AND CURVE.

§ 1. Ex. LXXVIII. *The Development of Algebraic Symbolism.*—The exercises reviewed in this chapter are not very closely connected but present a certain degree of unity through their relation to the theory of the logarithmic and antilogarithmic or exponential functions.

Ex. LXXVIII aims at a brief exposition of the view of the nature and development of algebraic symbolism which has already been set out at some length in the first chapter of this book. The teacher may, at his discretion, use the material of that chapter to amplify the discussion. The argument is illustrated (i) by a revision of the development of the exponential notation as it actually occurred in the course of Part I; (ii) by a sketch of an alternative mode of development which will be recognized as the one usually given in textbooks. Something is no doubt to be gained by showing that the evolution of algebraic ideas may proceed in more than one way; on the other hand, some teachers may think the advantage of a wider view to be neutralized by the risk of confusing the student by a double presentation. Truth to tell, if prudence did not suggest that the demands of examiners must not be ignored, the author would have chosen to emphasize the argument of Section III rather than to offer an alternative. Reasons for this attitude have been given already (pp. 57, 300); they may be supplemented by the contention that the method of Section III is much more direct and arithmetical than the usual method, and is therefore more in accordance with the most fruitful tendencies of modern mathematics. Consideration will show, in fact, that the review in Ex. LXXIII of the fundamental operations on numbers is incomplete through its omission of the operations

represented by the logarithmic and exponential notations; for the definition of the exponential operation as repeated multiplication (or division), like the definition of multiplication as repeated addition, breaks down when the operating number is not an integer. Thus the symbolism a^b presents a theoretical problem of exactly the same kind as those solved in the cases of the symbolisms $a + b$ and ab. In all three cases we have a pair of numbers (a, b) which is to be replaced by a single number c in accordance with certain principles of equivalence, and the problem is to find a general definition which will include as special cases the different principles used according as the numbers a and b are integral, rational, or irrational.

In the case of equivalences of the form $a^b = c$ the nearest approach we have yet made to a universal definition is the rule that if m and n are any values whatsoever of b, then a^m and a^n must be subject to the relation

$$a^m \times a^n = a^{m+n}.$$

In other words, the principle of equivalence which enables us to write $a^m = c_1$ and $a^n = c_2$ must also enable us to write $a^{m+n} = c_1 c_2$. Thus our task resolves itself into the search for a definition of the symbolism a^b from which this formal rule will follow in the case of *all* numerical values of a and b. The following definition (which is a paraphrase of the one given by Cantor[1]) will be found to possess the required generality. Let M and N be two classes or sets of elements whose numbers are respectively a and b. Then, as we saw in Ex. LXXII, ab may be defined, for all real values of a and b, as the number belonging to the collection of pairs which can be made by associating each element of M with each element of N. According to the definition now to be studied, a^b is to mean another mode of association between the elements of M and N—one in which we consider not pairs but sets of pairs made up upon the following plan. Take each of the elements of N and associate with it *any* one of the elements of M, in such a way that a given element of M may be either left unpaired or associated with 1, 2, 3, . . . or all of the elements of N. Let c be the number of the *sets* of pairs which can be formed in this way; then, by definition, $c = a^b$. As an illustration let P, Q and R be the elements

[1] See Russell, *Principles*, p. 308.

of N and p, q the elements of M. Then a^b is represented by scheme I:—

I.			II.	
P	Q	R	S	T
p	p	p	p	p
q	q	q	q	q
p	p	q	p	q
q	q	p	q	p
p	q	q		
q	p	p		
p	q	p		
q	p	q		

The rows below the top line exhibit the elements which are combined with P, Q and R to give the successive sets. The scheme represents the case in which $a = 2$, $b = 3$, while $a^b = 8$, as it should do in accordance with the usual definition. Similarly, scheme II shows the case in which $a = 2$ as before, $b = 2$ and $a^b = 4$.

In accordance with our definition scheme I represents a class of 2^3 elements, each element being one of the *rows* of letters. Similarly scheme II represents a class of 2^2 elements. If now we want to construct the class whose number is the product of 2^3 and 2^2 we must, by the definition of multiplication, associate each *row* of the one scheme with each *row* of the other. In this way a more elaborate scheme will be produced consisting of rows of letters associated with the five letters P, Q, R, S, T. The following arrangement shows a few of the rows:—

III.

P	Q	R	S	T
p	p	p	p	p
q	q	q	p	p
p	p	q	p	q
.	.	.	.	.
.	.	.	.	.

It is evident that in scheme III we have the two letters p, q associated with the five letters P, Q, R, S, T in the manner laid down in the definition under consideration. We conclude that the number of terms in the product-scheme may be represented as 2^5 and therefore that

$$2^3 \times 2^2 = 2^5.$$

Now it is clear in the first place that this argument could be repeated to prove that

$$a^m \times a^n = a^{m+n}$$

for all integral values of a, m, and n. It will further be seen that it establishes the same relation when the numbers are not necessarily integral but are any real numbers. For in this case the elements p, q, . . . P, Q, R, . . . and S, T, . . . may be taken to be the sets of rationals of which the real numbers $\mathbf{a}$, $\mathbf{m}$, and $\mathbf{n}$ consist, and it is evident that, although the infinite collections of elements now connoted by these symbols cannot be set down on paper, yet the schemes of association are in conception just as definite as before. That is to say, we can conceive (though we cannot picture) schemes of association in which the rationals contained in the real numbers $\mathbf{m}$ and $\mathbf{n}$ play respectively the parts assigned to P, Q, R and S, T in schemes I and II, while the rationals contained in the real number $\mathbf{a}$ take up the rôle of p and q. Further, we can conceive these modes of association combined into a scheme whose relation to the schemes just described is exactly that of III to I and II, the elements P, Q, R, S, T being replaced by the rationals contained in the real number $\mathbf{m} + \mathbf{n}$. We conclude that, the foregoing interpretation of the exponential notation being adopted, the relation

$$\mathbf{a^m \times a^n = a^{m+n}}$$

holds good for any real numbers $\mathbf{a}$, $\mathbf{m}$, and $\mathbf{n}$. Thus our new definition of the equivalence $a^b = c$ has universal validity; for it not only coincides with the original definition in the case when the symbols represent integers but also leads to the formal law with which *any* admissible enlargement of the original use of the notation must comply.

In Ex. LXXIII, p. 23, we saw that after arriving at a general definition of the equivalence $\mathbf{ab = c}$ we had still to face the practical problem of determining a value for $\mathbf{c}$ when $\mathbf{a}$ and $\mathbf{b}$ are given. Similarly, the foregoing definition of the equivalence $\mathbf{a^b = c}$ must be supplemented by a rule for assigning to $\mathbf{c}$ a numerical value to correspond to given values of $\mathbf{a}$ and $\mathbf{b}$. For this rule we must return to the method of ch. XXXIV., noting that, although the definition studied in the present article applies with equal precision to integral, rational, and irrational values of the variables, yet it is possible in the last two cases to calculate only approximately the actual arithmetical equivalences.

The teacher will observe that the foregoing argument carries the analysis much farther than it is taken in Ex. LXXVII, and he must decide for himself whether or not to supplement the simpler discussion offered there. If he decides upon the completer treatment he will find it convenient to introduce it in connexion with **No. 14** of the examples. For that example is meant to bring out the fact that the definition hitherto given of the functions

$$y = log_a x \text{ and } y = a^x$$

fails to secure complete generality. The proof is very simple. Referring to the scheme on p. 341 of this book we see that the logarithm as there defined is always a multiple of a rational number $k(= 1/p)$. Thus, in the function $y = log_a x$, y must always be rational, and in the function $y = a^x$, x must always be rational. In the alternative definition of Ex. LXXVIII, § 3, x, in the function $y = a^x$, is necessarily of the form p/q where p and q are integers; whence the same results follow again. Thus the antilogarithm of an irrational such as $\sqrt{2}$ (**No. 15**) can be defined only indirectly. In the upper line of the scheme on p. 341 there will be always one multiple of k which is less than $\sqrt{2}$ while the next multiple above it is greater than $\sqrt{2}$. In correspondence with this fact the antilogarithm of $\sqrt{2}$ will lie between a certain power of h and the next higher power. The numerical gaps between the two multiples and the two powers constantly decrease as p increases and may be made less than any assigned number by taking p sufficiently large. Thus antilog $\sqrt{2}$ may be defined as the number which always lies—no matter how large p may be—between the two powers of h which correspond to the two multiples of k between which $\sqrt{2}$ lies.

The teacher will see that these solutions of **Nos. 14, 15** are independent of the rather abstruse argument of the former part of this article, and are in themselves interesting and instructive.

§ 2. Exs. LXXIX-LXXXI. *Annuities. Life Insurance.*—From the purely theoretical work of the previous exercises we now turn to topics of an entirely practical nature. The teacher who hesitates to demand from his students the sustained abstract thinking required in Exs. LXX-LXXVIII may prefer to begin at this point. The subject of compound interest and annuities has already received some consideration

in Part I (Exs. XXXVI, C, LVII, B, LIX). The simple treatment there given is here expanded into a doctrine which deals with the essentials of the theory of private annuities, public loans, and life insurance. These subjects are of great interest from both the mathematical and the broader educational standpoints. It should be unnecessary to apologize for directing the student's attention to the scientific principles involved in matters of such immense private and public concern. The author has attempted to make the treatment as "real" as the necessity for brevity permits. The teacher should ensure the success of this attempt by supplementing the examples of the textbook by examples from actual life. A few well-chosen advertisements in the columns of the daily newspaper, the prospectus of a building society, and the tables of benefits and premiums of an insurance company may be used to secure, with an absurdly small expenditure of effort, a great return in interest and understanding and in enhanced respect for the social value of mathematics. The equipment of technical knowledge needed to give profitable instruction in this field is small and easily obtainable. *Whitaker's Almanack* or a similar publication is essential; most modern encyclopædias and many inexpensive "business handbooks" provide the necessary explanations and commentaries. The reader who seeks a deeper knowledge of the mathematical aspect of the subjects will find what he needs in King's *Theory of Finance* and the exhaustive *Text Book of the Institute of Actuaries*.[1]

§ 3. Ex. LXXXII. *The Exponential Function and Curve.*—In this exercise we resume, extend, and generalize the study of "growth-curves" and their corresponding functions which occupied our attention in Section III. In that section the growth-curve, whatever the growth-factor, was assumed always to be in the position in which the ordinate at the origin is unity. In this "standard position" it corresponds to the function $y = r^x$ where r is the growth-factor. We are now (i) to explore the functions which correspond to other positions of the curve, and (ii) to see that a growth-curve can always be represented as correspond-

[1] Part I of the *Text Book* deals with annuities-certain, public loans, etc.—ground covered in a simpler but sufficient manner in King's work. Part II deals with the theory of contingent annuities and life insurance.

ing to a function in which r is replaced by the "standard growth-factor" e.

In accordance with the principle established in chs. xxv. and xxvi. the graph of **No. 1**, shifted through $-a$ scale-divisions horizontally, corresponds to the function $y = r^{x+a}$. But we have

$$\begin{aligned} y &= r^{x+a} \\ &= r^a \cdot r^x \\ &= \mathrm{A}r^x \end{aligned}$$

where $\mathrm{A} = r^a$. This result is independently evident from the general property of the curve; if the ordinate at the origin is A the ordinate at distance x must be $\mathrm{A}r^x$. The inverse function takes (**No. 2**) two corresponding alternative forms

$$y = \log_r x \text{ and } y = \log_r(x/\mathrm{A}) = \log_r x - a.$$

The graph of the former inverse function is obtained from that of $y = r^x$ by the usual two steps of (i) turning the original graph in the plane of the paper through a clockwise right angle, and (ii) revolving it about the x-axis through 180°. A further lowering through a distance a (corresponding to the leftward shifting of the original graph) produces the graph of the second inverse function. These principles are applied in **No. 3** and the following examples. Thus in **No. 3** (i) we have

$$\begin{aligned} y &= 3(10)^x \\ &= (10)^{0\cdot477}(10)^x \\ &= (10)^{x+0\cdot477} \end{aligned}$$

the substitution being made, of course, by reference to the table of logarithms. We conclude that the graph of the given function is that of $(10)^x$ shifted (approximately) 0·477 scale-units to the left. Similarly in (iv) we have

$$\begin{aligned} y &= \log_{10} 27x \\ &= \log x + \log 27 \\ &= \log x + 1\cdot431. \end{aligned}$$

That is, the graph of $(10)^x$ has first been rotated in its own plane, then out of its plane about the x-axis, and then raised a distance 1·431. In **No. 4** (iii) the graph, after shifting, corresponds to

$$\begin{aligned} y &= (3\cdot2)^{x+4} - 13 \\ &= 104\cdot86\ (3\cdot2)^x - 13. \end{aligned}$$

In **No. 10** we have

$$\begin{aligned} y &= \mathrm{A}r^x \\ &= \mathrm{A}(e^p)^x \\ &= \mathrm{A}e^{px} \end{aligned}$$

where $e^p = r$. Applied in **No. 11** (ii) this result gives

$$y = -2(0{\cdot}8)^x$$
$$= -2e^{-0{\cdot}223x}$$

since $\log_e 0{\cdot}8 = \log_e (1/1{\cdot}25) = 0 - 0{\cdot}2231$ by the table of natural logarithms.

The examples in division B are important not only on account of their physical applications but also (and especially) in view of the argument of the next exercise. They do not, however, offer any particular difficulty. For **No. 17** (which may be omitted) reference may be made to pp. 43-5 of this book.

§ 4. Ex. LXXXIII. *Differential Formulæ.*—This exercise deals with the differential formulæ of the exponential and logarithmic functions, and is very important both for that reason and because it gives a good occasion for revising the whole doctrine of differential formulæ. The main argument is based upon the properties of the exponential curve which were brought to light in ch. xxxv. Even if the class has worked the examples (Ex. LIX) which followed that chapter it will no doubt be advisable to revise its conclusions before passing to the present exercise. It will probably be well also to revise the earlier work on differential formulæ (ch. xxvii., B).

Nos. 1-4 prepare the way for the main argument by establishing certain properties of the "first differences" of the ordinates of the exponential curve. **No. 1** is quite simple but of fundamental importance. We have

$$\frac{\mathrm{P'Q'}}{y} = \frac{ar^{x+h} - ar^x}{ar^x}$$
$$= r^h - 1$$

which is obviously constant so long as h, the distance between the ordinates, is constant. The geometrical argument of the note before **No. 5** shows that when h is so small that it can be treated as a differential then the constant $r^h - 1$ assumes the value $\log_e r \,.\, \delta x$. Since at the same time P'Q' becomes δy the result of No. 1 is now the differential formula

$$\frac{\delta y}{y} = \log_e r \,.\, \delta x \qquad \text{or} \qquad \frac{\delta y}{\delta x} = ar^x \,.\, \log_e r.$$

When $r = e$ (**No. 6**) $\log r = 1$, and this equivalence reduces to the familiar result that when $y = ae^x$

$$\frac{\delta y}{\delta x} = ae^x.$$

The remaining examples of division A are simple applications of this extremely important relation. **No. 11** needs a reference to *Exercises*, I, p. 261. The solution of **No. 12** is as follows:—

Since $$\frac{\delta^2 y}{\delta x^2} = 10e^x$$

it follows that $$\frac{\delta y}{\delta x} = 10e^x + b.$$

To find the value of the undetermined constant b we note that, when $x = 0$, $\delta y/\delta x = \tan 42° = 0{\cdot}9$ nearly. Hence

$$0{\cdot}9 = 10 + b$$

or $$b = -9{\cdot}1$$

and $$\frac{\delta y}{\delta x} = 10e^x - 9{\cdot}1.$$

From this relation again we deduce that

$$y = 10e^x - 9{\cdot}1x + a$$

where a is a second constant to be determined by the consideration that $y = +12{\cdot}3$ when $x = 0$. Making the substitutions we have

$$+12{\cdot}3 = 10 + a$$

or $$a = +2{\cdot}3.$$

Whence finally $$y = 10e^x - 9{\cdot}1x + 2{\cdot}3.$$

With **No. 13** we turn to the differential formula of the logarithmic function. Since the function $y = \log_r x$ is the inverse of $y = r^x$, it follows that the differential formula of the former is the inverse of that of the latter. That is to say, to find the differential formula of $\log_r x$ we have only to interchange x and y in the result of No. 5. When this interchange is effected we have

$$\frac{\delta x}{x} = \log_e r \,.\, \delta y$$

or (since $\log_e r = 1/\log_r e$)

$$\frac{\delta y}{\delta x} = \frac{\log_r e}{x}.$$

In **Nos. 16, 17** this result is used to demonstrate in a more satisfactory and comprehensive way than in Part I, Ex. LXIX the universal validity of Wallis's Law. In **No. 17** we have (since $y = x^n$) $\log y = n \log x$, logarithms being taken

to base e. But if we put $z = \log y = n \log x$ we have (as in No. 16)

$$\frac{\delta z}{\delta y} = \frac{1}{y} \text{ and } \frac{\delta z}{\delta x} = \frac{n}{x}$$

whence, dividing the second result by the first, we have

$$\begin{aligned} \frac{\delta y}{\delta x} &= n\frac{y}{x} \\ &= n\frac{x^n}{x} \\ &= nx^{n-1}. \end{aligned}$$

The examples in divisions B and C are all important and should be carefully worked and clearly understood. A weakness in this part of the work will be a source of much trouble at many subsequent points, and pains should be taken to avoid it. **No. 26** is of great historical importance and will be used more than once in the sequel. The answer to the last part of the question is, of course, that there is no logarithm of zero; thus when zero is one of the values of x the rule for the hyperbolic area breaks down. In other words, there is, strictly speaking, no "area-function" corresponding to the "ordinate-function" $y = a/x$; that is, there is no function which gives the whole area under the curve right from the y-axis. The area may, however, be calculated from an ordinate as near as we please to the y-axis provided that it is not actually coincident with or on the other side of it. This same result, in modified forms, recurs in **Nos. 29, 30.** We have previously met it (Part I, Ex. LXIX, No. 14) in the form of the statement that Wallis's Law breaks down when $n = 0$.

The examples of **No. 30** imply a knowledge that the differential formula of the function $y = \log_e (x + a)$ is

$$\frac{\delta y}{\delta x} = \frac{1}{x + a}.$$

The simplest proof is to consider the curve $y = \log_e x$ and to note that the gradient at the point where the abscissa is x is given by $\delta y/\delta x = 1/x$. Let the curve be shifted horizontally to the left through a distance a. Then the number previously denoted by x must now be symbolized by $x + a$. Thus the ordinate-function becomes $y = \log_e (x + a)$ and the gradient-function $\delta y/\delta x = 1/(x + a)$.

In division C the relation between a differential formula and its primitive is employed to calculate the "volume-function" of a solid whose parallel cross-sections exemplify a definite "area-function". These examples are not only important as further illustrations of the power of the method in mensuration; in addition they are valuable as preparing for the general ideas of a "derived function" and "integral" which are to be reached in Section VIII.

The exercise ends with a few examples in which differential formulæ are used to solve simple problems in kinematics. These examples are to be regarded as a continuation of those of Part I, Exs. XXX, B, and LXIX, C. In connexion with them the teacher should read again the remarks on p. 171 of this book and the note on p. 353 of *Exercises*, Part I.

§ 5. Ex. LXXXIV. *Supplementary Examples.*—Revision Papers 1-4 contain nothing but examples of types already studied in Part I and require no commentary. Revision Paper 5 (division E) presents in a simple way the essentials of the theory of "scales of notation". These examples may be regarded as a continuation of those of Part I, Ex. XXXI, C.

Division F may be regarded as supplementary to Ex. LXXXIII. In it the student is led, following the steps of Lord Brouncker, Wallis, Mercator and Gregory, to a knowledge of some of the most important "series" in elementary mathematics. These results were, historically, achievements of the highest significance, not because they made it possible to calculate logarithms by processes easier than those of Napier and Briggs—for the work of calculating the tables was already accomplished—but because they showed the immense potentialities of Wallis's method and opened altogether new vistas in mathematics. The reader who would understand how suggestive these pioneer investigations proved should seek access to Francis Maseres' *Scriptores Logarithmici*, a monumental work in six large volumes, published in the last half of the eighteenth century and containing reprints (and sometimes translations) of all the more important memoirs which had their origin in the wonderfully fruitful work of Napier and John Wallis.

SECTION V.

TRIGONOMETRY OF THE SPHERE.

THE EXERCISES OF SECTION V.

*** The numbers in ordinary type refer to the pages in *Exercises in Algebra*, Part II; those in heavy type to the pages of this book.

EXERCISE		PAGES
LXXXV.	MAP PROJECTIONS; SANSON'S NET . .	105, 442
LXXXVI.	CYLINDRICAL PROJECTIONS . . .	110, 443
LXXXVII.	MERCATOR SAILING	117, 445
LXXXVIII.	GREAT CIRCLE SAILING	124, 447
LXXXIX.	CALCULATIONS ON GREAT CIRCLE SAILING .	133, 454
XC.	SOME ASTRONOMICAL PROBLEMS . .	143, 457
XCI.	SUPPLEMENTARY EXAMPLES	
	A. Stereographic projections . .	157, 449
	B. Projections in general . . .	162, 452
	C. Astronomical problems . . .	167, 462
	D. Supplemental triangles, etc. . .	169, 466

CHAPTER XLII.

PROJECTIONS.

§ 1. *Projections.*—The work of the section begins with four exercises on the theory and use of map projections. These serve a triple purpose. First they bring the student into close quarters with the laws governing the spatial relations of points upon a spherical surface. There is, in fact, no equally effective means of bringing out the striking differences between these laws and those of plane geometry. Secondly, they offer, probably, by far the simplest mode of approach to the trigonometrical calculations by which problems of position on a sphere are solved. Thirdly, they are excellent concrete instances of the general notion of a projection or "transformation"—a notion which must be considered as one of the fundamental ideas of mathematics.

Speaking generally, a projection [1] is any rule or device by which, given a set of points A, B, C, D, etc. in a line, surface or volume, we can obtain a second set A′, B′, C′, D′, etc. in another line, surface or volume, corresponding to the first set, point by point. A plan (e.g. of a field) drawn to scale is a simple example of a projection, for the essence of plan-making is that to every point in the original there shall correspond a definite point in the drawing. In the case of a plan the whole of the spatial relations of the original points are reproduced; only the metric scale is changed. In projection in general the relations may all be transformed; it is sufficient if they are transformed in any regular way which secures point-to-point correspondence between the original and the representation.

The cartographer's problem is one of projection, for he has to represent in a definite way upon a flat sheet of paper the

[1] The term is here used to cover both projections proper and other transformations. See *Exercises*, II, p. 162.

spatial relations of points on a sphere. In representing them he is bound to transform them. This fact makes the fundamental difference between a map and a plan. But it is possible so to choose the mode of representation that some given feature of the original space-relations is preserved. Thus for the geographer it is important that a map should represent correctly the relative sizes of land and sea surfaces. The cartographer furnishes him, therefore, with an "equal area projection". The sailor demands a chart which shall facilitate the task of navigation. He receives, therefore, either a "Mercator," a "gnomonic" or a "stereographic" projection. The point to make clear in teaching is that each of these projections preserves something and that the choice of a projection is determined by the purpose for which it is to be used.

It should be added that many of the maps actually used by geographers are not projections at all, in the strict sense, for the positions of the representative points are not determined by a single law. Different laws are used in different parts of the map. The explanation of this usage is simple but should be given in the geography lesson. In the mathematics lesson attention should be concentrated upon projections which have mathematical as well as practical value. The most important from this point of view are those dealt with in the exercises. For further information about them and about the theory of map-projections in general, the teacher may consult the following works: *Ency. Brit.*, art. "Maps"; Hinks, *Map Projections*.

Finally the teacher is strongly urged to make a free use of the globe in these lessons. A "blackboard surface" globe is most useful. The geographical equipment of a secondary school frequently includes a number of small globes of this kind which can be put into the hands of individual students.

§ 2. Ex. LXXXV. *Sanson's Net.*—Sanson's "sinusoidal" net is chosen for the first exercise as a simple instance of an equal area projection. The completed drawing is represented in the frontispiece of *Exercises*, II (firm lines). Its reproduction is not too laborious an exercise for an evening's homework, but may be lightened by the omission of alternate meridians and parallels. When the curves are satisfactorily drawn in pencil they should be redrawn in ink so that the

pencil lines used in solving problems may be rubbed out without damage to the net.

The solution of **No. 2** is obvious. Between a given meridian and the central meridian $l/360$ of each parallel is intercepted, l being the difference of longitude between the meridians. But (by p. 131) the length of a parallel is $2\pi r \cos \lambda$ where λ is the latitude or $2\pi r \sin p$ if p is the distance in degrees from the north or south pole (i.e. $p = 90° - \lambda$). Hence the intercept of the parallel whose polar distance is p is

$$\frac{\pi r l}{180} \sin p.$$

That is, the meridian is a sine curve whose "amplitude" is $\pi r l/180$.

The circles of **Nos. 6, 7, 8** are represented in the frontispiece of *Exercises*, II (firm lines). It is well to bring out the point of the problems by cutting from a sheet of card a circular hole whose radius is that of the circles in question. That is, the diameter must be the chord between two points on the equator of the globe separated by 45°—easily measured by a pair of compasses or dividers. When the card is laid on the globe with the centre of the hole at the prescribed point the circumference of the hole passes through the specified positions on the globe in each case.

§ 3. Ex. LXXXVI. *Lambert's Net.*—The importance (for mathematics) of Lambert's projection is that it leads to a simple formula for the area of any belt of a sphere and therefore to the formula $A = 4\pi r^2$ for the whole sphere. The argument in § 1 contains a point worthy of special emphasis, since it involves the principle upon which the whole treatment of the calculus (Section VIII) is to be founded. It is assumed (i) that the area of a sphere lies between the area of any system of conical surfaces which can be inscribed in it and any similar system in which it can be inscribed; (ii) that as h decreases the areas of the inscribed and circumscribed surfaces continually approach the area of the sphere but can never coincide with it. It is shown that corresponding to each value of h there is (i) a cylinder, as high as the sphere but of smaller radius, whose surface is equal to that of the inscribed conical system; (ii) and another cylinder, also of equal height with the sphere but of larger radius, whose surface is equal to that of the circumscribing conical system. It is also shown that

as h decreases these cylinders continually approach the cylinder which circumscribes the sphere but can never coincide with it. Thus we have two series :—

. . . E, E, E, . . . S . . . I, I, I, I, . . .

. . . *e*, *e*, *e*, . . . C . . . *i*, *i*, *i*, *i*, . . .

—a series of external and internal conical surfaces (E, I) and external and internal cylinders (*e*, *i*). The Es are separated from the Is by the sphere S; the *e*s from the *i*s by the circumscribing cylinder C. The members of the respective series approach S and C endlessly. That is why no member can be represented as lying next to S or C; it would always be possible to imagine another nearer still. But to every E an *e* corresponds and to every I an *i*. It follows, therefore, that S which divides the Es from the Is must correspond to C which divides the *e*s from the *i*s. That is to say, the area of the sphere is identical with that of the circumscribing cylinder whose height is equal to the sphere's diameter.

It will be observed that this method of proof leads to *exact* results. It is, therefore, greatly preferable to the proofs which depend upon the assumption that a narrow belt of the sphere can itself be regarded as a conical frustum. The student always feels such a proof unsatisfactory. It is true that the error, as regards a single belt, may be diminished without limit by diminishing its breadth, but the number of the belts are magnified in the same proportion. The total effect of the accumulation of errors remains, therefore, uncertain. In any case such a method gives the area of a system of conical surfaces, not of the sphere.

The teacher who can command the use of skilled fingers will find it useful to construct *Exercises*, II, figs. 67 and 68, in wire. Rotated by a turning table, by a gyroscopic top or simply by the fingers, the wire frame will give a cinematographic representation of the sphere with the system of conical surfaces and the corresponding cylinder.

Lambert's net is represented by the broken lines of the frontispiece of *Exercises*, II. The same figure also contains the solution of **No. 5.**

Section B of this exercise gives an interesting anticipation of the results of "differentiating" and "integrating" the sine and cosine of an angle. A class which has not worked through Part I will need an explanation of the terms "area-function" and "ordinate-function". See ch. XXVII., A.

The examples on the "central cylindrical" net serve a two-fold purpose. First, the net offers a useful contrast to Lambert's. In the latter the distance of a parallel from the equator is $r \sin \lambda$ and can, therefore, never exceed r. In the former (**No. 13** (ii)) it is $r \tan \lambda$ and therefore increases endlessly as the latitude increases. Thus it is theoretically impossible to show the north and south poles in a central cylindrical projection, and practically inconvenient to show much beyond the Arctic and Antarctic Circles. Secondly, the net is our first example of a "geometrical projection" (Ex. XCI, § 4). It is generally possible to purchase from an ironmonger a wire frame of the kind used to protect naked gas flames in workshops and theatres. The wires often give a passable imitation of the lines of latitude and longitude on a globe. By means of a lighted candle with its flame at the centre the shadows of the wire may actually be cast on to a paper cylinder held round the frame.

§ 4. Ex. LXXXVII. *Mercator Sailing.*—In this exercise we resume the problems in navigation studied in Exs. XIX and XX.[1] When a sailor has in contemplation a voyage of some hundreds of miles he cannot regard the surface of the sea as a plane. It becomes essential, therefore, to have a method by which to guide his ship from port to port. It will be seen that there are two methods in actual use. The first ("Mercator sailing") aims at finding the "rhumb line" which will carry the sailor from port to port without any change of direction. That is the subject of Ex. LXXXVII. The second method aims at determining the *shortest* track between the two ports. That is studied in the next exercise. The student is usually greatly surprised to find that the "straight" course is not also the shortest. It is well to keep the discovery back until Ex. LXXXVII is finished.

The teacher will note in § 3 of this exercise a repetition in essence of the argument emphasized above on p. 444. He will also recognize that the graphic method of § 3 is a device for obtaining a practical solution of the integral $\int \sec \theta \, . \, \delta\theta$. A theoretical solution is given at a later stage (Ex. CX, D). By following the directions the class should arrive at the following :—

[1] It is not necessary to have worked these exercises in order to proceed with the present one.

TABLE OF MERIDIONAL PARTS.[1]

Lat.	5°	10°	15°	20°	25°	30°	35°	40°
Pts.	5·01	10·05	15·18	20·42	25·83	31·47	37·40	43·71
Lat.	45°	50°	55°	60°	65°	70°	[75°]	[80°]
Pts.	50·50	57·91	66·13	75·46	86·31	99·43	[116·17]	[139·6]

The examples in division A of the exercise can be solved by

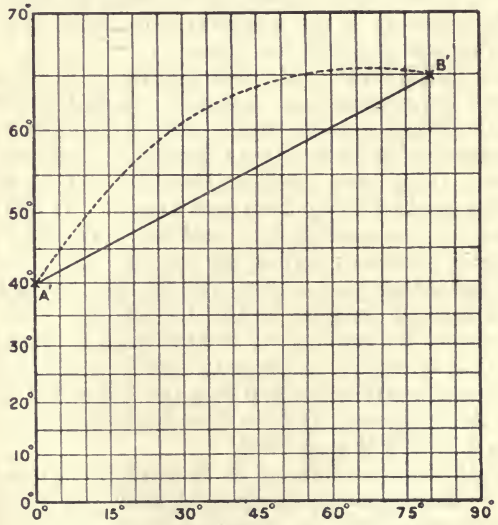


FIG. 101.

a Mercator net which contains 180° of longitude and 70° of latitude on one side of the equator. One half of such a net is represented in fig. 101. In different problems the left-hand

[1] The unit is the length of a *degree* along the equator. In the larger tables (e.g. Chambers') the unit is the length of a *minute* along the equator.

edge of the net must be taken to represent different meridians. In dealing with places south of the equator the net must be inverted.

Taking the left-hand edge to be the meridian of 100° W. the straight line A′B′ represents the rhumb line of Ex. LXXXVIII, **No. 6.** The curved line represents the great circle track joining the same two points. The angle between A′B′ and the meridian through A′ is the course which would take a ship from A′ to B′ without change of direction. Call it α, and let l be the difference of longitude between A′ and B′ —that is the length of the parallel through B′ measured in equator-degree units. Finally let m_1 and m_2 be meridional parts corresponding to the latitudes of A and B. That is let m_1 and m_2 be the distances in equator-degree units of A′ and B′ from the base line of the net. Then we have

$$\tan \alpha = l/(m_2 - m_1).$$

This is the solution of **No. 12.**

To calculate the distance D between A and B we suppose the track to be broken up into pieces so short that the whole of a piece may be supposed to have the same latitude. If d be the length of such a piece in miles then $d \cos \alpha$ gives the north and south distance in miles between its two ends. Since α is the same for each piece D cos α is the distance in miles between the parallels of latitude through A and B. But each mile is a minute of a degree. Hence the difference of latitude in degrees is D cos α/60. Thus, given the difference of latitude and the course, the rhumb-line distance can at once be found (**No. 14**).

§ 5. Ex. LXXXVIII. *Great Circle Sailing.*—This exercise is of great importance because the spherical trigonometry of the next two exercises is based upon it. It should, however, give no difficulty to teacher or student. The wire frame of § 3 above serves usefully as a means of demonstrating both the polar and meridian gnomonic nets. Fig. 102 shows (on a reduced scale) the polar gnomonic net of **No. 2.** In **No. 3** the pole and the parallels are supposed to be south of the equator, in **No. 5** north. The straight line A′B′ in fig. 102 gives the solution of **No. 5**, the curved line that of **No. 7**, while the lines joining A′B′ in fig. 101 are the solution of **No. 6.**

The argument of § 3 of the exercise is usefully supplemented by a model which can be constructed with little difficulty. A

child's india-rubber ball about 17 cms. in diameter is blackened. A figure BAP consisting of a great circle track and two meridian segments is drawn on it in white paint. The sphere is placed in contact with an upright sheet of glass, the line OP being horizontal. *Exercises*, II, fig. 72, is drawn on stiff card, the arcs PA, AB, BP, having the radius of the ball. The card is cut along the arcs and cut half through along the lines A′P, PB′, B′A′. The three truncated triangles are bent down and bound together with tape along their edges AA′ and BB′, and are placed so that the arcs PA, AB, AP, rest on the

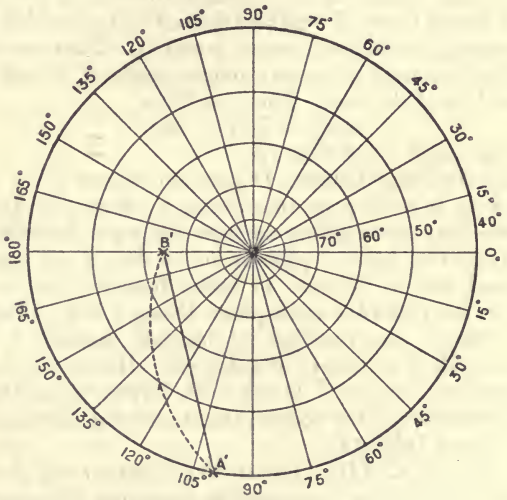


FIG. 102.

painted arcs on the ball, while PA′B′ rests against the glass. Finally, narrow strips of paper are stuck on the other side of the glass covering the edges of the cardboard. The mechanism of the projection can in this way be made clear to a class. The cardboard can be removed, leaving the triangle on the ball and its projection on the glass. Indeed the cardboard itself, without the ball and glass, makes a very useful model, since it exhibits at the same time the spherical triangle and its plane projection. Whether this model is used or not, each student should make (and keep for future use) the simple paper model of **No. 9.**

To solve **No. 12** mark the positions of Honolulu (A′) and Yokohama (B′) on the polar gnomonic net, transfer the triangle PA′B′ to a sheet of paper and complete the construction of *Exercises*, II, fig. 72. Bisect the angle A′OB′ by OC′ cutting A′B′ in C′. Mark the corresponding point C′ in the gnomonic net and read its latitude and longitude.

Fig. 103 shows the meridian gnomonic net of **No. 13**. The straight line across it gives the solution of **No. 14**. To meet

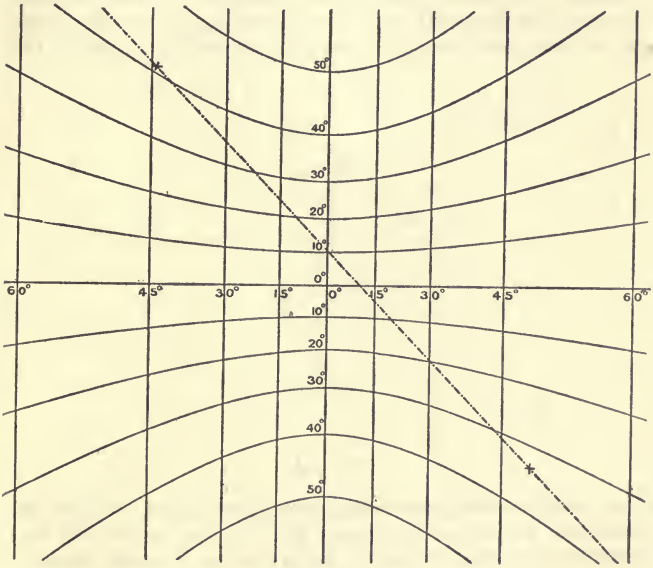


FIG. 103.

the requirements of the problem the central meridian, though marked zero, is taken to be the meridian of 30° W. If it had been taken (for example) as the meridian of 40° W. or 25° W. the actual position of the line would have been different in the diagram, but the answers to the questions set would have been the same. This statement should be verified.

Fig. 104 gives the diagram described in § 5. It is important in connexion with the formulæ of right-angled spherical triangles (Ex. LXXXIX, **No. 25**).

§ 6. *Supplementary Work : The Stereographic Projection.*—

To division A of the supplementary Ex. XCI are relegated two more projections. Of these the stereographic will well repay study if the necessary time can be afforded. It has two beautiful properties: (i) any circle on the sphere is projected as a circle, and (ii) the angle at which the projections of any two curves cross one another is the same as the angle at which the actual curves cross upon the surface of the sphere. The arguments required to establish these truths give excellent work for the student. Taken together with the fact that it is easy to construct a net for the greater part of the globe they also make the projection a useful one

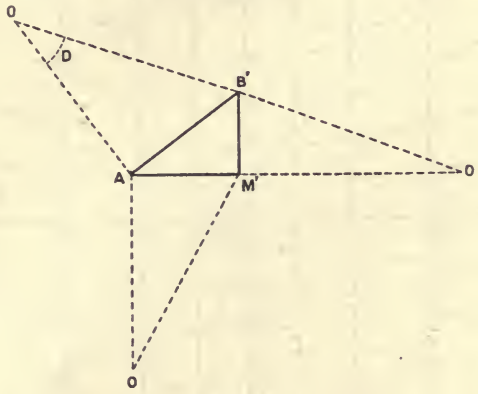


Fig. 104.

in connexion with great circle sailing problems. Unlike the gnomonic projection it shows the correct course at any moment of a ship which is sailing along a great circle.

Fig. 105 shows the net described in **No. 4** together with the projections of the three circles mentioned in the examples. It will be seen that they are also circles.

The stereographic projection also gives the easiest means of studying the sum of the angles of a spherical triangle. It is easy to use it to demonstrate that the sum must always exceed two right angles (**No. 11**). It is left to the teacher to point out that the larger the sphere the greater will be the diameter of the great circles on it, for sphere and great circles have the same radius. Thus if, in the fig. of No. 11, AB′, AC′ and the angle B′AC′ remain of constant size but the

sphere becomes larger, the diameter B′B″ will increase and the angle B′B″C′ (which is half the spherical excess) will diminish and approach zero. On the other hand, if AB′ and AC′ increase, since the diameter of the great circle remains constant, the spherical excess increases. We may conclude, then, that the larger the sphere or the smaller the

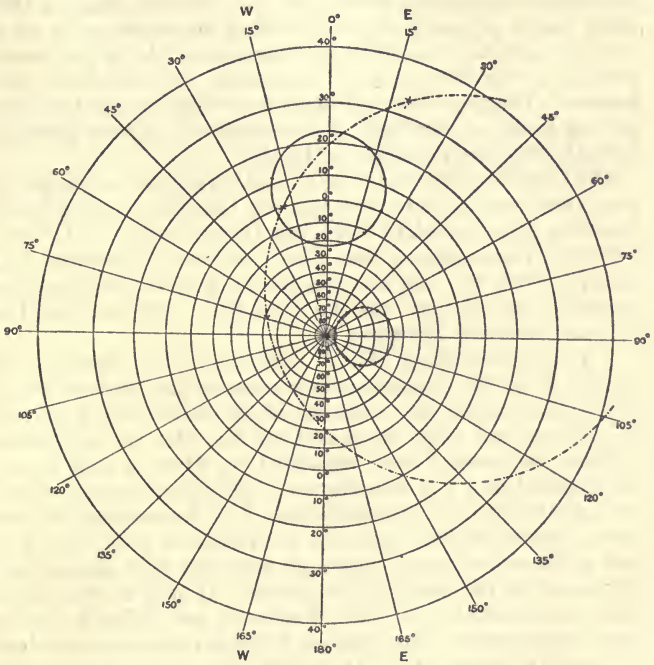


FIG. 105.

triangle the smaller is the spherical excess; and that the smaller the sphere or the larger the triangle the greater the spherical excess.

The teacher may, if he feels disposed, refer to the speculation that points in what we think of as plane surfaces *may* have really the spatial relations which characterize the surface of a large sphere. Measurement can inform us only that the sum of the angles of a triangle is approximately 180°; it

cannot decide whether apparent deviations from that sum are due to defective observation or a real curvature in space. The only way to settle the question would be to measure the angles of an extremely large triangle (such as that marked out by the centres of three fixed stars) in which the spherical excess (if it exists) might be expected to show itself. The speculation can be continued further. We can suppose that space has a property to be described metaphorically as a *negative curvative* the effect of which would be to give every triangle a "spherical defect" increasing in magnitude with its size. This hypothesis is also one which cannot be tested so long as we are confined to the relatively minute triangles of earth, but may nevertheless be true.

The teacher who finds himself—and his pupils—attracted by such ideas is advised to refer to Prof. Carslaw's translation of Bonola's *Non-Euclidian Geometry* (Open Court Publ. Co.); Clifford's *Philosophy of the Pure Sciences* ("Lectures and Essays," Vol. I); Stallo's *Concepts of Modern Physics* (Int. Scientific Series) may also be consulted. The subject has now an enormous literature.

§ 7. *Supplementary Work; Projections in General.*—In division B of the supplementary exercise the student is led first to summarize the properties which he has found in map projections and then to generalize the idea of projection. Nothing substantial need be added to what is said in the exercise and in § 1 of this chapter. Reference may, however, be made to the brief discussion (in § 7 of the exercise) of one-to-one correspondence between the points of a straight line and a plane—a notion which is, perhaps, best approached by means of the study of projection. It will be recognized that the elaboration of this idea is far too difficult for an elementary course; nevertheless it is convenient to introduce it if only to show that it *is* difficult and must, therefore, be treated with respect. Moreover, taken with Ex. LXXI, the discussion helps to illuminate the difficult but important idea that there is a multiplicity of "infinite," just as there is of "finite," numbers, each possessed of its own individuality and properties (see p. 408).

The argument relies too much upon geometrical intuition to be really satisfactory, but it may awaken interest in the question.

The interesting point which emerges from the discussion

is that although a surface contains an "infinite" number of separate lines, each possessing an "infinite" number of points, yet the points in any surface and in any line must be considered to have the same number, for they can be brought into one-to-one correspondence. The teacher who wishes to follow the subject further should turn to Young's *Fundamental Concepts* (Lect. XVI) or Russell's *Principles of Mathematics*. Further brief references to it will also be found in Ch. LII and in Ex. CXVI.[1]

[1] The following title should be added to the works recommended at the end of § 1: Mary Adams, *A Little Book on Map-Projection* (Geo. Philip & Co.). The teacher will find this book (which appeared too late for mention in the text) extremely lucid and informative.

CHAPTER XLIII.

THE TRIGONOMETRY OF SPHERICAL TRIANGLES.

§ 1. *The Programme.*—Exs. LXXXIX and XC investigate the application of trigonometrical formulæ to problems concerning the relations of points upon the surface of a sphere. The problems which are of practical importance here fall into two classes. In the first are those of the navigator and the surveyor who have to deal with points upon an actual spherical surface. In the second are the problems of the astronomer who finds it convenient to state his problems in the form of problems about points upon a sphere, though the sphere is only imaginary, or at most a model representing in an easily intuitable form his observations upon the stars. A few additional problems of almost purely geometrical interest are relegated to the supplementary exercise (division D) and may, without serious loss, be omitted.

It has already been remarked in ch. XXXVIII., § 3 that the technical apparatus needed for mastering everything of first-rate importance in this field is quite small in extent. It does not go beyond the fundamental formula of spherical triangles :—

$$\cos a = \cos b \cos c + \sin b \sin c \cos \mathrm{A} \quad . \quad (1)$$

the three formulæ of right-angled triangles :—

$$\sin \mathrm{A} = \sin a / \sin c \quad . \quad . \quad . \quad (2)$$

$$\cos \mathrm{A} = \tan b / \tan c \quad . \quad . \quad . \quad (3)$$

$$\tan \mathrm{A} = \tan a / \sin b \quad . \quad . \quad . \quad (4)$$

and the derived formula :—

$$\frac{\sin a}{\sin \mathrm{A}} = \frac{\sin b}{\sin \mathrm{B}} = \frac{\sin c}{\sin \mathrm{C}} \quad . \quad . \quad . \quad (5)$$

The "supplemental formulæ" corresponding to these are given (as a cheap luxury) in the supplementary exercise, but "Napier's analogies," and other complications introduced in order to obtain formulæ adapted to logarithmic computation are entirely beyond the requirements of the ordinary student

and are therefore omitted altogether. The classification of problems into six "cases" (including the "ambiguous case" with its formidable array of sub-divisions) is also excluded as an enterprise from which the ordinary student is likely to gain very little profit.

§ 2. Ex. LXXXIX. *Navigation Problems.*—Division A of this exercise is devoted to establishing the fundamental formula (1) of § 1. The formula is reached by considering a concrete problem in navigation—namely to find the great circle distance between places whose latitudes and longitude difference are given. This problem has already been solved graphically in Ex. LXXXVIII. The solution by formula demands nothing more than the application of simple trigonometry to *Exercises*, II, fig. 72. The student is told (**No. 1**) to take this figure (or the paper model which he made in Ex. LXXXVIII, No. 9) and to enter against each side an expression for its length in terms of the data.

In the triangle OA′P it is obvious that $OP = r$, $PA' = r \tan p_1$ and $OA' = r \sec p_1$. But OA′ in this triangle is identical with OA′ in the triangle OA′B′, for when the two triangles are folded respectively about PA′ and A′B′ the similarly lettered edges fall together. Thus in the triangle A′OB′ we have $OA' = r \sec p_1$, and in the same way $OB' = r \sec p_2$. In the triangle PA′B′ we can now write:—

$$(A'B')^2 = r^2 \tan^2 p_1 + r^2 \tan^2 p_2 - 2r^2 \tan p_1 \tan p_2 \cos l$$

and in the triangle OA′B′

$$(A'B')^2 = r^2 \sec^2 p_1 + r^2 \sec^2 p_2 - 2r^2 \sec p_1 \sec p_2 \cos D.$$

Equating these expressions, cancelling the r^2 throughout, and remembering that $\sec^2 a - \tan^2 a = 1$, we reach the first equivalence of **No. 4** from which the formula for cos D follows at once.

In the first instance this formula is proved only for cases which can be represented by *Exercises*, II, fig. 72—that is, when the track AB does not cross the equator. In § 2 of the exercise the formula is without difficulty shown to hold good equally when this restriction is removed.

So far the student has dealt only with concrete problems about great circle distances, etc. In division C, § 3, he is first taught to see that the methods used in these problems are applicable whenever we are dealing with a figure composed of the arcs of three great circles, and that the fundamental formula can be applied, under proper conditions, to deter-

mine any side or angle of such a figure. At this point the term "spherical triangle" is first introduced. The idea of the angle between two "sides" of a spherical triangle needs careful attention. The paper model can here be used again with useful effect.

In § 4 the important special case of the right-angled spherical triangle is considered. The method of treatment depends upon the fact that the gnomonic projection can, in this case, be made to preserve unchanged both the right angle *and* one of the other angles of the figure on the sphere. Thus, when C is the right angle and the plane of the projection makes contact with the sphere at A, we have $AB' = r \tan c$, $AC' = r \tan b$ and (since the angle A is conserved)

$$\cos A = \tan b / \tan c.$$

If, in this projection, AB′ and AC′ are regarded (like PA′ and PB′ in *Exercises*, II, fig. 72) as *meridians* we have no direct information about the length of B′C′ and can, therefore, derive no expressions for sin A and tan A. If, however (as suggested in **No. 25**), A be regarded as the crossing point of the central meridian and the equator of fig. 103, AC′ as a part of the *equator* and B′C′ as a meridian, we have a means of supplying values to all the sides in the projection, and (since the angles A and C are again conserved) can at once obtain expressions for all the trigonometrical ratios of A. The figure required is, of course, fig. 104.

In division D these formulæ are used to solve various problems with regard to great circle paths on a sphere. Most of these are couched in the form of questions in which the latitude, longitude and course of a ship are to be expressed with reference to the "vertex" of its track—that is, the point where the track comes nearest to the pole or where the perpendicular great circle from the pole meets it. The results of these problems are generalized and summarized in **No. 35**. The formulæ indicated in this example play, with regard to a great circle, the same part as the formula

$$y = ax + b$$

plays with regard to a straight line in a plane. That is, they enable us to determine whether a given point is or is not situated upon a given great circle. The variety of the formulæ which serve this purpose is due to the fact that the position of a point on a great circle may be fixed by any two out of three coordinates: namely, its latitude, its longitude, and its

distance from the vertex. In **No. 35** (i) the coordinates to be used are the co-latitude and longitude. We are supposed to know AB ($= p$), BC ($=$ P) and the angle at C ($= 90°$). The angle at B is L $-$ l. Hence by formula (3) (p. 454) we have

$$\cos (\mathrm{L} - l) = \tan \mathrm{P}/\tan p,$$

a formula which may be regarded as stating the relation between the values of the variables l and p for all points along the great circle whose vertex is at (L, P). Similarly in **No. 35** (iv) we know BC ($=$ P), AC ($= d$) and the angle C ($= 90°$) and have to find a formula for AB ($= p$). In this case we use the fundamental formula (1) which reduces to

$$\cos p = \cos \mathrm{P} \cos d$$

—an expression which states the relation between p and d for all points in a great circle whose vertex is in latitude $90° -$ P.

Lastly, these formulæ make it possible to determine the direction of a great circle at any point—that is the angle at which it crosses one of a given system of meridians. Thus in **No. 35** (v) we are to determine A ($= a$), given BC ($=$ P), AB ($= p$) and C ($= 90°$). By formula (2) (p. 454) the required relation is

$$\sin a = \sin \mathrm{P}/\sin p.$$

§ 3. Ex. XC. *Astronomical Problems.*—The first two examples are intended to suggest the use of gnomonic projections in recording the positions of the stars at a given moment. The lines inscribed upon the roof of the imaginary cubical room will form a polar gnomonic net in which lines of equal azimuth take the place of the meridians of Ex. LXXXVIII, and circles of equal altitude the place of the parallels of latitude, while the pole of the former net becomes the zenith. The walls will be inscribed with four meridian gnomonic nets in which the horizon takes the place of the equator. The complete net of **No. 2** is represented in fig. 106.

If the cubical room is tipped about its northern edge through an angle equal to the co-latitude of the place of observation (§ 3, **No. 3**) the azimuth lines become lines of equal right ascension or equal hour-angle, the altitude lines become lines of equal declination, the horizon becomes the "celestial equator" and the zenith the "celestial pole". If "stars" are stuck on to the glass so as to represent the complexion of the sky when the "first point of Aries" is on the meridian (i.e. when the sidereal clock registers 0 h. 0 m.

0 s.), the room must be re-oriented to represent the sky at another hour (**No. 4**). Thus when the sidereal clock registers 3 h. the roof must have revolved in its own plane through an angle of 45° (= 15° × 3). It should be noted that since the sidereal clock gains about 4 m. a day upon the solar clock the interval measured in units of solar time will be 3 h. − 4 m. × $\frac{3}{24}$ = 2 h. 59 m. 30 s. For naked-eye observations this difference is, of course, negligible.

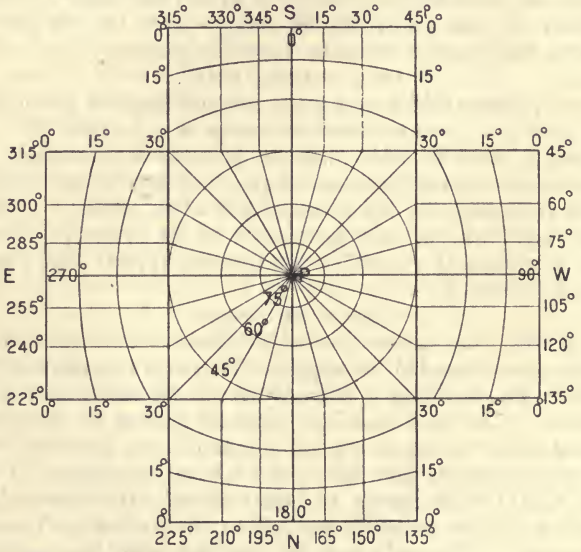


Fig. 106.

A revolving cubical room is obviously a practical impossibility, but it is easy to gain most of the advantages it has to offer by enlarging the scope of the altitude circles inscribed on the roof and suppressing the sides. We thus obtain the perfectly practicable star-net of **Nos. 5** and **6**. Such a net should certainly be made and mounted in the way described in No. 6. The positions of a number of prominent stars should be entered upon it.[1] The card should be held (**No. 6**)

[1] The following is a satisfactory minimum list: the Pole Star, the seven bright stars of the Plough, the "W" of Cassiopeia, the stars Vega and Capella. The R.A. and Decl. of these stars may be

so that the line joining its centre to the eye is parallel to the axis of the sky's rotation and has a length equal to the radius assumed in constructing the net. The R.A. line whose graduation is identical with the momentary reading of the sidereal clock should point from the pole of the card towards the zenith. The "stars" on the card will then be identical in position with the stars in the sky.[1] The position of the horizon can be indicated permanently by a slight addition to the apparatus. Fix a strip of cardboard or paper across the base in such a way that the circular chart can turn round beneath it. Arrange that its upper edge (where it is a chord of the circular chart) shall touch the circle whose polar distance is equal to the latitude (or whose declination is equal to the co-latitude). If the chart is now held in the position prescribed above and with the edge of the paper strip horizontal, the latter as viewed by the eye will coincide with the horizon. Any "star" which is below the edge of the strip will be invisible. It is possible in this way to determine the sidereal times of the rising and setting of any star which is represented on the chart.

Nos. 9 to 14 illustrate the way in which observations of the altitude of a star on crossing the meridian can be used to determine latitude. They prepare the way for the study of the way in which this same determination is constantly made at sea by observations on the meridian altitude of the sun.

§ 4. *The use of Trigonometrical Formulæ.*—In § 4 the astronomical globe is introduced as a convenient means of representing the observed facts of stellar movement. It then becomes obvious that the formulæ of spherical triangles can be used to solve numerous problems by calculation. The problems here considered almost all resolve themselves into

read from any star-map—for example from those contained in Ball's *Popular Guide to the Heavens* (Geo. Philip & Son). *Whitaker's Almanack* also gives the positions of most of them.

[1] To find the sidereal time at a given moment take from *Whitaker's Almanack* the sidereal time at noon on the day in question and add the time of the observation. Thus on 2 May, 1913, the sidereal time at noon was 2 h. 39 m. 10 s. The sidereal time at 7.30 p.m. was, therefore, practically 10 h. 10 m. The difference between sidereal and solar units is here ignored since only approximate results are needed.

the determination of the altitude and azimuth of a star whose hour-angle and declination are given—or into the converse problem. The analogy of the problem of **No. 15** with that of calculating a great circle distance should be made clear—preferably with the help of a globe. The zenith distance may be regarded as the great circle distance between the zenith and the star. The other two sides of the spherical triangle are the polar distances of the zenith and the star—the former being, of course, a constant for a given place of observation. The hour-angle takes the place of the difference of longitude between the two ends of the great circle track. The method of determining the azimuth is explained in **No. 16.**

The solution of **No. 21** is as follows. As usual the formula should be *remembered* in the form (1) but *used* in form (2) :—

$$a = 54° 23', \quad \delta = 17° 35' 36'', \lambda = 42° 18'$$

$$\cos h = \frac{\sin a - \sin \delta \sin \lambda}{\cos \delta \cos \lambda}$$

$$= \frac{2 \sin a - \cos(\lambda + \delta) + \cos(\lambda - \delta)}{\cos(\lambda + \delta) + \cos(\lambda - \delta)}$$

$$= 0{\cdot}8645$$

$$\therefore h = 30° 11'$$

$$= 2 \text{ h. } 0 \text{ m. } 44 \text{ s.}$$

That is, the observation took place 2 h. 0 m. 44 s. after Saturn crossed the meridian. Since this passage occurred at 3 h. 41 m. 27 s. by the sidereal clock the sidereal time of the observation must have been 5 h. 42 m. 11 s. But the sidereal time at noon was 20 h. 20 m. 41 s. The observation was, therefore, made

24 h. + 5 h. 42 m. 11 s. − 20 h. 20 m. 41 s. = 9 h. 21 m. 30 s. after noon. The time is measured here in sidereal units and is practically $9\frac{1}{3}$ sidereal hours. To reduce it to solar hours we must deduct

$$4 \text{ m.} \times \frac{9\frac{1}{3}}{24} = 1 \text{ m. } 33 \text{ s.}$$

We conclude that (within a few seconds) the local time was 9.20 p.m.

§ 5. *Calculations based upon Solar Observations.*—Calculations based upon observations of the sun are complicated by the irregularity of the daily movements of that body. For this reason they are postponed until the student has

mastered the simpler problems connected with stellar position. The main facts about the sun's movements in right ascension are best studied in connexion with the sundial; consequently division C of the Exercise begins with the problem of graduating a dial when the base upon which the shadow falls is horizontal.[1] This is, of course, not the simplest form of dial. The simplest possible form consists of a style (e.g. a hat-pin) thrust through a square of cardboard or wood at right angles to its plane. Upon this the shadow of the style is to be received. The lower end of the style must be driven into a wooden horizontal base at an angle of inclination equal to the latitude. The edge of the receiving square should rest on the same base. The angle which this square makes with the base will, of course, be the co-latitude. If the dial is placed so that the style is parallel with the axis of the sky's rotation it will be found that at all seasons of the year the shadow of the style moves uniformly over the receiving surface at the rate of 15° per hour. This is, perhaps, the best way to introduce the young observer to the knowledge of the law of the sun's rotation. The calculations of Nos. 25-27 presuppose this law and are based upon it.

Whatever form of sundial is employed the main points which emerge from a comparison of its readings with those of a good clock are those covered by the term "equation of time". In accordance with the principle ch. xxxviii., § 3, they are treated in § 5 of the exercise simply as facts of observation. They are not difficult to explain but the explanation lies in the province of the text-book on astronomy. Our present purpose is limited to applying them in simple calculations of the same type as those considered in the earlier examples.

The solution of the first part of **Nos. 31** and **32** is as follows:—

Since the longitude of Bristol is 2° 35′ W., the local time corresponding to $7.10\frac{1}{3}$ a.m. Greenwich time is 7 a.m. At 7 a.m. on 28 May the sun is

$$3 \text{ m.} + 7 \text{ s.} \times \tfrac{5}{24} = 3 \text{ m. } 1{\cdot}5 \text{ s.}$$

before the clock. It is therefore 4 h. 56 m. $58\frac{1}{2}$ s. before noon —the time being measured in solar units. In accordance

[1] The article *Dialling* in the *Ency. Brit.* is very full and instructive. There are also special books on the subject.

with the explanation which precedes No. 31 the hour-angle of the sun is the angular equivalent of this interval, namely 74° 14′ 39″. The sun's declination at 7 a.m. was

$$21° 24′ 41″ - 24·5″ \times 5 = 21° 22′ 38·5″$$

To find the altitude (a) we have the formula

$$\begin{aligned} \sin a &= \sin \delta \sin \lambda + \cos \delta \cos \lambda \cos h \\ &= 0·44263 \\ a &= 26° 16′ 18″ \end{aligned}$$

To find the azimuth we apply the law of sines (No. 16) in the form :—

$$\sin \beta = \sin h \,.\, \cos \delta / \cos a$$

whence

$$\beta = 88° 3′$$

The exercise concludes with examples (**Nos. 35** and **36**) on the method by which the latitude and the local mean time are determined simultaneously by observing the altitude of the sun as it crosses the meridian. This is the standard method of nautical astronomy. The longitude is determined by the difference between the local mean time and the Greenwich mean time recorded by the ship's chronometer.

In former days the occasional determination of Greenwich mean time by astronomical observations was itself one of the ordinary tasks of the navigator. Such observations are now required much less frequently—the rapid progress of wireless telegraphy have made them to a great extent unnecessary. At the present day "wireless" time-signals are flashed out daily from suitably placed stations all over the world and are picked up and transmitted from ship to ship along all the great trade-routes. A uniform system of signals was adopted at an International Conference held in Paris in 1912. An interesting account of it is given in *Nature* (London) for 20 March, 1913. An account of the "standard time zones" accompanied by a map will be found in Ball's *Popular Guide to the Heavens*.

§ 6. Ex. XCI, C. *Supplementary Examples.*—Supplementary examples on spherical triangles are given in divisions C and D of Ex. XCI. In division C those on "lunar distances" need no additional explanation. **Nos. 33** and **34** give point to the account of the sun's movements in the previous exercise by showing that time can be read system-

atically by means of the shadow of a vertical rod. The angle β which the shadow makes with the north and south line (**No. 33**) is, of course, equal to the sun's azimuth. The sun's altitude must first be calculated by the formula

$$\sin \alpha = \sin \delta \sin \lambda + \cos \delta \cos \lambda \cos h$$

and the value of β derived from that of α by the formula

$$\sin \beta = \sin h \,.\, \cos \delta / \cos \alpha.$$

Taking the latitude of London as $51\frac{1}{2}°$ the highest possible altitude when the hour-angle is $15° \times 3 = 45°$ is given by

$$\begin{aligned}\sin \alpha &= \sin 23° 27' \,.\, \sin 51\tfrac{1}{2}° + \cos 23° 27' \,.\, \cos 51\tfrac{1}{2} \,.\, \cos 45° \\ &= 0{\cdot}71528 \\ &= \sin 45° 40'.\end{aligned}$$

In finding the lowest possible altitude in the same circumstances we must remember that $\sin \delta$ was, for convenience, substituted for $\cos P$ in the original spherical formula. When the sun's declination is 23° 27′ *south*, $p = 90° + 23° 27'$ and $\cos p = -\sin 23° 27'$. Thus we have

$$\begin{aligned}\sin \alpha &= -\sin 23° 27' \,.\, \sin 51\tfrac{1}{2}° + \cos 23° 27' \,.\, \cos 51\tfrac{1}{2}° \,.\, \cos 45° \\ &= 0{\cdot}09238 \\ &= \sin 5° 18'.\end{aligned}$$

Substituting the two values of α successively in the second formula we have

$$\sin \beta = \frac{\sin 45° \,.\, \cos 23° 27'}{\cos 45° 40'}$$

$$\beta = 68° 10'$$

for the first, and

$$\sin \beta = \frac{\sin 45° \,.\, \cos 23° 27'}{\cos 5° 18'}$$

$$\beta = 40° 39'$$

for the second. We conclude that during the course of the year the direction of the shadow 3 hrs. before (or after) noon will vary in London to the extent of 27° 30′.

Nos. 35 and **36** deal with the fascinating problem of graduating a sundial which is to be erected upon a vertical wall. There are two cases: (i) when the plane of the wall lies east and west (No. 35); (ii) when it is inclined to the east and west line (No. 36).

The solution of the first of these problems will be aided by consideration of figs. 107 and 108. In these figures the northern edge PQ of a horizontal dial is supposed to be

pressed against the bottom edge PQ of a vertical dial, and the dials are supposed to have a common style OO′. Fig. 107 shows a meridian section of the two dials which passes through the style. If l is the length of the style it is seen from this figure that ON = $l \cos \lambda$ and O′N = $l \sin \lambda$. In fig. 108 the vertical dial is supposed turned about PQ into the same plane with the horizontal dial. OA is the shadow on the latter when the hour-angle is h. It is evident that the shadow on the vertical dial must pass through O′ and also through A; its position is, therefore, the line O′A. Let the

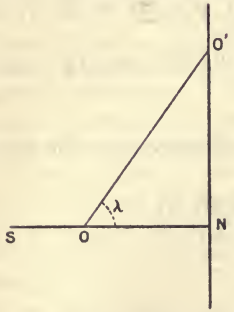


Fig. 107.

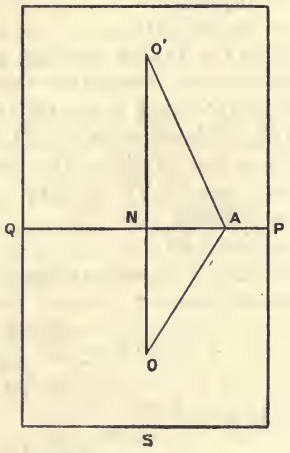


Fig. 108.

angle NOA = H and the angle NO′A = H′. Then we have

$$ON \cdot \tan H = NA = O'N \cdot \tan H'$$

i.e. $$\tan H' = \tan H \times \frac{ON}{O'N}$$

$$= \tan H \cdot \cot \lambda.$$

But by Ex. XC, No. 25, $\tan H = \sin \lambda \cdot \tan h$; hence

$$\tan H' = \cos \lambda \cdot \tan h.$$

Next let the wall be inclined at an angle of $w°$ to the EW line (**No. 36**). In order that the line ON on the horizontal dial may still be in the meridian when the edges of the dials are brought together its base must be cut, as in fig. 109, along the line PQ so that the angle PNE = $w°$. When the dial,

thus modified, is pressed against the wall fig. 107 still holds good; that is to say, the style still makes an angle $90° - \lambda$ with the vertical line O'N in the plane of the meridian. But the plane of the meridian instead of being perpendicular to the wall now makes an angle of $90° - w$ with it. This consideration determines the position of the style with respect to the wall.

To determine the graduations we suppose, as before, that the vertical dial is brought into one plane with the horizontal

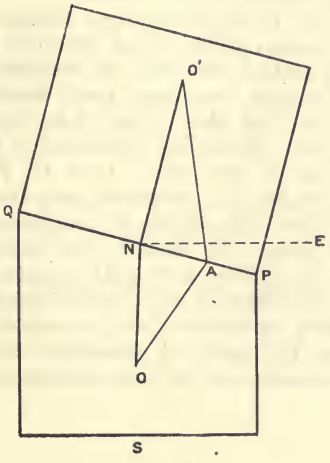


FIG. 109.

dial (fig. 109). To calculate NA in the triangle ONA we observe that

$$\angle NAO = 180° - H - \angle ANO$$
$$= 90° - H + w.$$

Whence we have

$$\frac{NA}{\sin H} = \frac{ON}{\sin NAO}$$
$$= \frac{ON}{\cos (H - w)}$$

and

$$NA = ON \times \frac{\sin H}{\cos (H - w)}.$$

But, in the triangle O'NA, NA = O'N . tan H' as before. Hence

$$\tan H' = \frac{ON}{O'N} \times \frac{\sin H}{\cos (H - w)}$$
$$= \frac{\sin H \,.\, \cot \lambda}{\cos (H - w)}.$$

It is not possible this time to make a simple substitution for H in terms of h. The hourly values of H must be calculated separately and inserted in the formula.

§ 7. Ex. XCI, D. *Supplementary Examples.*—Ex. XCI, D, deals with certain properties of spherical triangles which are of great theoretical but only of secondary practical importance. The most interesting point about the area of a spherical triangle (§ 10, **Nos. 37-42**) is its connexion with the "spherical excess" previously considered in **Nos. 11** and **12** and discussed in ch. XLII., § 6. In § 11, **Nos. 43-50**, the student is led to see the reciprocal connexions which exist between the sides and angles of a spherical triangle and the angles and sides of the related polar triangle, and to use these relations to deduce additional formulæ for the solution of problems. To what is said in the exercise nothing need be added except a repetition of the recommendation that demonstrations on the spherical blackboard should accompany the teacher's exposition of all such subjects as this.

SECTION VI.

COMPLEX NUMBERS.

THE EXERCISES OF SECTION VI.

*** The numbers in ordinary type refer to the pages of *Exercises in Algebra*, Part II ; the numbers in heavy type to the pages of this book.

EXERCISE		PAGES
XCII.	The Nature of Complex Numbers . .	175, 474
XCIII.	Products of Complex Numbers . . .	181, 475
XCIV.	Complex Values of the Independent Variable	186, 478
XCV.	Complex Values of a Function . .	191, 481
XCVI.	The Relations between two Complex Variables	201, 486
XCVII.	The Logarithm of a Complex Number .	205, 493
XCVIII.	Supplementary Examples	
	A. The "Exponential Values" of the sine and cosine	213, 494
	B. Circular functions of the complex variable	214, 494

CHAPTER XLIV.

THE NATURE OF COMPLEX NUMBERS.

§ 1. *The History of "Imaginary" Numbers.*—It has already been pointed out that the term "imaginary numbers" is (like the term "irrational numbers") simply a relic of a past in which the real significance of these mathematical entities had not yet been perceived. It was first used by Descartes in his *Géométrie* (1637) in connexion with the solution of equations. It is instructive to observe that even to so modern a writer positive roots are the only "true" or "real" solutions, negative roots being regarded as "false" on the ground that they claim to represent numbers which are less than nothing.[1] By the process (which Descartes discovered) of raising the values of the roots of equations it is possible to turn these "false" into "real" roots. There are, however, some roots which cannot be made "real" by increase or diminution; these Descartes, so to speak, abandoned as imaginary.[2] Thus the equation

$$x^3 - 6x^2 + 13x - 10 = 0$$

has only one "real" root, namely 2; the other two roots are the "imaginary" numbers $2 + \sqrt{(-1)}$ and $2 - \sqrt{(-1)}$.

The mathematical logic of the seventeenth and eighteenth centuries did not reach any satisfactory view of the nature of expressions of the form $a + b\sqrt{(-1)}$. From the time of De Moivre (*c.* 1730) onwards such expressions were used by mathematicians in intermediate stages of their arguments and calculations, but there seems always to have been a lingering doubt of the validity of results reached by such mysterious means. This is the reason why textbooks of trigonometry still supplement investigations in which imaginaries are employed by "proofs not involving the use of

[1] *Géométrie*, Bk. III, p. 78 in the edition of 1664.
[2] *Ibid.*, p. 86.

$\sqrt{(-1)}$". Through the connexion which Descartes established between algebra and geometry "imaginary" points, lines, etc., gradually entered into the latter science, until in modern times these strange terms have come to connote some of its most important conceptions.

The long delayed rational interpretation of "imaginary numbers" appeared almost simultaneously in three distinct quarters at the beginning of the nineteenth century. The Dane, Caspar Wessel, has the honour of priority, his tract *On the Representation of Direction* having been published at Copenhagen in 1799, seven years before the obscure Genevan, J. R. Argand, printed in Paris his now famous *Essai sur une manière de representer les quantités imaginaires dans les constructions géométriques.* Lastly, it seems probable that earlier than either of these writers the great Gauss had reached the views which, published in 1831, became the actual source of the modern doctrine of complex numbers.

Argand's essay, though it seems to have been little known until it was republished in 1874,[1] contains a wonderfully clear and confident exposition of the true doctrine of imaginaries, and is still worth careful study. The author starts from the sound philosophical position that an entity symbolized by $a + b\sqrt{(-1)}$ is not necessarily more "absurd" or "imaginary" than one symbolized by a negative number. Thus the result symbolized by $b - a$ when b is less than a is certainly "imaginary" unless we add to the conception of magnitude, which necessarily belongs to it as a number, the further conception of direction. When numbers are thus regarded as directed, operations and results which were previously impossible at once become admissible. For example, we can prolong any given arithmetic progression indefinitely far both ways, the repeated subtractions which carry the terms past zero being representable as equal steps taken along an endless straight line. May it not be that the so-called imaginary numbers simply carry a stage further the process by which the negative number is produced (by addition of the idea of direction) from the number without sign?

To answer this question Argand bids us consider the sequence

$$+1 \qquad x \qquad -1$$

[1] An English translation by Prof. A. S. Hardy was published in 1881 in Van Nostand's Science Series.

The intention here is that -1 shall bear the same relation to x as x bears to $+1$. That is, the operation which turns $+1$ into x must also be supposed to turn x into -1. Remembering that $+1$ and -1 must be representable as points, such as A and I in fig. 110, equally distant from a zero-point K, we see that any operation which satisfies the condition laid down must be representable geometrically by an operation which, in two identical applications, would carry a point from A to I. One such geometrical operation would be the movement of the point along AI through the distance AK. This is the operation which would generate an arithmetic sequence and is to be represented algebraically as the addition of -1. It would obviously give to x the value zero. Two other possible geometrical operations would consist in revolving the line KA through a right angle either in the direction of E or in the direction of N; for each of these operations when repeated would bring the point A into coincidence with I. Further double applications of the same operation would in each case bring the point successively back to A ($+1$), to I (-1), to A again, and so on for ever. Now there are two *algebraic* operations which, by definition, must produce, when applied in this same way, the sequence $+1, -1, +1, -1, \ldots$ for ever. They are multiplication by $+\sqrt{(-1)}$ and by $-\sqrt{(-1)}$. Thus for exactly the same reason that we identify -1 with a unit step taken along a line in a certain direction, we may identify $+\sqrt{(-1)}$ with a revolution of a line through a right angle in one sense and $-\sqrt{(-1)}$ with an equal revolution in the opposite sense. The choice of senses being arbitrary we take $+\sqrt{(-1)}$ to imply anticlockwise rotation.[1]

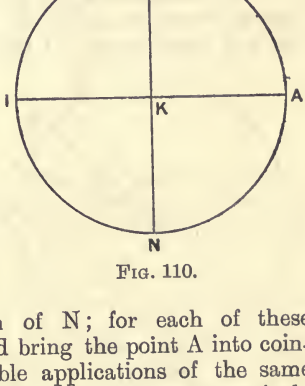


FIG. 110.

The remainder of Argand's essay is devoted to the ex-

[1] The reader should be warned that although the above description represents correctly the substance of Argand's argument the expression has been somewhat expanded.

pansion of this cardinal idea and to practical applications of it. The first step in the generalization of the method will be to inquire how to represent lines such as KP, KQ, KB (fig. 111) of which the first two divide equally the angle AKB. The preceding argument has indicated that the operation which carries the line KA from its original position to KP, and then to KQ and KB in equal swings must be represented as multiplication by a constant factor. It is therefore easily seen that the factor must be of the form which De Moivre introduced into algebra—namely, $\cos a + \sqrt{-1} \cdot \sin a$ where a is the constant angle between the rays drawn from K. The argument is as follows. If the length of KA is not unity but r, if there are $m - 1$ intermediate rays, and if the angle

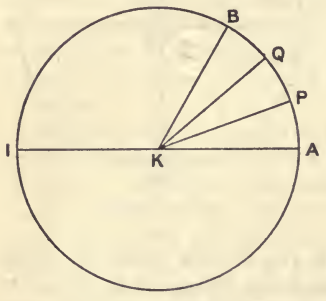


FIG. 111.

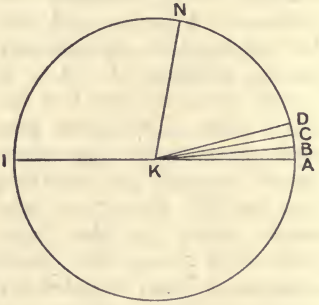


FIG. 112.

AKB is θ $(= ma)$, then the line KB can be represented by the expression $r[\cos\theta + \sqrt{(-1)} \cdot \sin\theta]$ or $a + b\sqrt{(-1)}$ where $a = r \cos\theta$ and $b = r \sin\theta$; for we know already (by De Moivre's theorem) that $\cos ma + \sqrt{-1} \cdot \sin ma = (\cos a + \sqrt{-1} \cdot \sin a)^m$. Argand called KB a "directed line" and spoke of r as its "modulus". The line of descent from the "directed line" to the "vector" and thence to Hamilton's "quaternions" is easy to trace.

It may be interesting to reproduce one of the applications which Argand makes of his theory, and we choose his derivation of the expansion for $\log(1 + x)$. In fig. 112 let the n arcs AB, BC, CD, etc., be equal. Then the "directed lines" KA, KB, KC, . . . KN, are to be identified with numbers of the form $a + \sqrt{-1} \cdot b$ which form a geometrical progression. Since the arcs AB, AC, AD, . . . AN are in

arithmetic progression they may (by Napier's definition) be taken as the logarithms of the terms of the geometric progression KA, KB, KC, . . . KN. Thus we may write

$$\log \mathrm{KN} = m\mathrm{AN}$$
$$= mn\mathrm{AB}$$

where m is an arbitrary constant whose value determines the "base" of the logarithms. Now let mn become great without end so that AB may be itself considered as a "directed line" at right angles to KA. Then we have

$$\mathrm{AB} = \mathrm{AK} + \mathrm{KB}$$
$$= -\mathrm{KA} + \mathrm{KB}$$
$$= -1 + (\mathrm{KN})^{1/n}.$$

Lastly, put $\mathrm{KN} = 1 + x$ and it follows that

$$\log(1 + x) = mn\{-1 + (1 + x)^{1/n}\}$$
$$= m\left(x - \frac{x^2}{2} + \frac{x^3}{3} - \ldots\right).$$

Gauss first published his views upon imaginary numbers in the "second commentary" upon his *Theoria Residuorum Biquadraticorum* (1831).[1] Many of his statements recall those of Argand. Thus he says that if $+1$, -1 and $\sqrt{(-1)}$, instead of being called positive, negative and imaginary units, had from the first been called direct, inverse and lateral units the confusion which darkened algebraic doctrine would never have arisen. Equally striking is his declaration that "the arithmetic of complex numbers is capable of complete intuitable representation [*anschaulichsten Versinnlichung*]". But the term "complex number" which he introduces in this discussion, shows that Gauss had reached a clearer analysis than Argand of the logical standing of the expression $a + \sqrt{-1}.b$. To him it implies simply a "couple" of real numbers, just as a rational number is a "couple" of integers; the radical difference between the new "complex numbers" and the old ones being that while the latter denote the positions of points upon a line the former denote the positions of points in a plane. In virtue of these geometrical applications we may say that real numbers constitute a one-dimensional series while complex numbers form a two-dimensional series.

When complex numbers had ceased to be a mystery and

[1] *Werke*, Vol. II.

could be viewed in the clear light of common sense, it began to be seen that $a + b\sqrt{(-1)}$ is really the typical number of algebra, and that "real" numbers should be regarded as merely special cases in which $b = 0$. The argument is as follows. Let $y = f(x)$ denote any function of x. Then if we are confined to real values of the variables we must admit that in the case of most functions there are either values of x to which no values of y correspond or values of y which are not produced by any value of x. But if our variables are complex numbers, these exceptions never occur; to a value of x of the form $u + v\sqrt{(-1)}$ there corresponds, in the case of every possible function, a value of y of the form $\mathrm{U} + \mathrm{V}\sqrt{(-1)}$, u, v, U, V, being themselves real numbers. Thus if we understand the word "number" in algebra always to mean "complex number," we can say with Cayley that "numbers form a universe complete in itself, such that, starting in it, we are never led out of it".

These observations have an obvious geometrical intrepretation. Real numbers correspond to points on a straight line, complex numbers to points in a plane. If we represent the values of x by points on one line and those of y by points on another we cannot say that every function $y = f(x)$ establishes one-to-one correspondence between all the points of the two lines; in most cases whole stretches of points will remain outside the correspondence. But if we take two planes, and represent the values of x by the points of one of them and the values of y by the points of the other, then one-to-one correspondence between *all* the points of both planes is effected by *every* function. The perception of this remarkable consequence of the nature of complex numbers found its clearest expression in Riemann's graphic method of representing functions which is to be studied in Ex. XCVII.

§ 2. Ex. XCII. *The Nature of Complex Numbers.*—The considerations brought out by the foregoing historical sketch are made the basis of the treatment in Ex. XCII and the following exercises. There is much to be said for beginning with quadratic equations and showing that, just as their "real" roots correspond to points on a line, so their "imaginary" roots correspond to points in a plane. But study of the exposition given in Exs. XCII, XCIII, will, it is hoped, show that a method which starts off boldly from Gauss's conception of the complex number as simply a couple or pair of real

numbers, the coordinates of a point in a plane, is really simpler and more interesting as it is certainly more in accord with the spirit of modern mathematics. Thus in Ex. XCII the form $a + ib$ is introduced as a convenient substitute for the notation (a, b), which has hitherto been used to indicate the position of a point in a plane. In this form the i is not a number but is merely a symbol warning us that the measurement b is to be taken at right angles to the direction in which a is measured. That is, i is Gauss's "lateral unit". After practice has been given in the use of complex numbers to represent points whose rectangular or polar coordinates are given, we turn to the problems of adding and subtracting them. To "add" two complex numbers $a_1 + ib_1$ and $a_2 + ib_2$ is found to mean: to take the movement which would bring a point from the origin to the point $(a_1\ b_1)$ and to follow it by a movement parallel and equal to the one which would carry a point from the origin to (a_2, b_2). It is also found that the solution of the problem can be reached with extreme ease if we treat the "lateral unit" i just as if it were a number. It is vital that the student should realize that this usage is entirely arbitrary and is adopted purely for the sake of the resulting convenience of manipulation. The teacher will see that the convention is made much more patent by being introduced before i has been identified with the quasi-numerical entity $\sqrt{(-1)}$.

§ 3. Ex. XCIII. *Products of Complex Numbers.*—This identification is the aim of the next exercise. The argument is simple but the teacher should study it with care since (for the reasons already given) it reverses the usual order of treatment. We begin by observing that $\cos\alpha + i\sin\alpha$ may be regarded as what Hankel[1] called a "direction coefficient," that is, a complex number which, when it multiplies another number, produces a result which corresponds to the turning of a line through the angle α. If we multiply the result by another factor of the same form, $\cos\beta + i\sin\beta$, it is reasonable to suppose that the new product describes the original line rotated through an angle $\alpha + \beta$. Scrutiny shows that this result would follow from mere algebraic manipulation of the expressions if, in addition to the assumption that i can be treated as a number, we assume that it can be treated as if its

[1] *Theorie der Complexen Zahlen-Systeme.*

square were -1. As is pointed out (*Exercises*, II, p. 182) it is not surprising to find that i is not to be regarded as equivalent to any number already recognized as such. If it were so its use could not correspond to a totally new idea. If the "lateral unit" is to be represented numerically at all the symbol must be one not hitherto thought to represent a number, just as the representation of the "inverse unit" involved symbolism which had not before been regarded as corresponding to any possible number.

After this first identification of the lateral unit i with $\sqrt{(-1)}$ it is necessary to make sure that the equivalence holds good under the same conditions in all cases. This task is carried out in the form of an investigation which establishes "De Moivre's Theorem" for all rational exponents. If it is needed to prove it also for irrational exponents the argument used on p. 432 to deal with irrational logarithms can, with suitable modification, be reapplied.

§ 4. *Complex Numbers and Vector Algebra.*—Any treatment of complex numbers which follows up the fruitful suggestions of Argand is bound to lead into the algebra of vectors. Informal excursions into this fascinating province have been taken at various points of the course from Part I, Ex. XIX, onwards. The author's conviction is that a few lessons in vector algebra form an excellent top storey to the edifice of school mathematics; for since the laws of this science depart from those of the algebra of numbers—even when the numbers are complex—the study of them greatly illuminates the principles of ordinary algebra. After some hesitation, however, he decided that the inclusion of a section on the algebra of vectors would be too serious an addition to the programme of this book. He must be contented, therefore, with mentioning a few works which deal suitably with the subject. Clerk Maxwell's *Matter and Motion* and Clifford's *Dynamic* are classical introductions to it, and the former can be read without difficulty by beginners. Henrici and Turner in their *Vectors and Rotors* have developed mainly the geometrical aspect, and Hayward in *The Algebra of Coplanar Vectors* the algebraic side. Kelland and Tait's well-known *Introduction to Quaternions* is a clear and useful textbook for those who wish to make acquaintance with Hamilton's powerful methods. Two books, recently published, will be found of the greatest use, G. Goodwill's *Elementary*

Mechanics (Clarendon Press) to those who have little or no acquaintance with mechanics or physics, and L. Silberstein's *Vectorial Mechanics* (Macmillan) to those who have already some knowledge of those subjects and are able, therefore, to appreciate the extraordinary powers of the vectorial calculus.

CHAPTER XLV.

RELATIONS BETWEEN A REAL AND A COMPLEX VARIABLE.

§ 1. *The Aim of Exs. XCIV, XCV.*—Let $y = f(x)$ be any function of x, and let account be taken, in the first place, only of real values of the variables. Then, as we have seen, those values can be represented by points upon two straight lines, and their connexion by a system of correspondences between the points. If the two lines are set at right angles to one another as "axes" the correspondences determine a "graph" of the familiar character, i.e. a straight or curved line lying in the plane of the axes. Next let account be taken of complex as well as of real values of x and y. Then in this case, since the values of each variable require a plane for their representation, there can be no graph in the ordinary sense. All that can be done is to exhibit in some way the correspondence which the function brings about between the points of the two planes. Between these two extreme cases come (i) those in which, while account is taken only of real values of y, complex as well as real values of x are considered, and (ii) those in which, while complex values of y are admitted, all except real values of x are excluded. In such cases one of the variables is a one-dimensioned number and the other a two-dimensioned number; it is possible, therefore, to represent their connexion by a graphic line drawn in three-dimensioned space. In the former class of cases values of y will be represented by points on the y-axis, values of x by the points of a plane; and since the two dimensions of this plane must be different from that already occupied by the y-axis it must be the plane through the x-axis perpendicular to the paper. Similarly in the second case the values of x may be represented by points upon the x-axis and the values of y by points in the plane through the y-axis at right angles to the paper.

Exs. XCIV and XCV are devoted successively to the study of these two types of three-dimensioned graph and of the functional connexions which correspond to them. The study is important for other reasons besides the fact that it simplifies the transition from the ordinary graph to the Riemann method of exhibiting the relations between the variables when both are complex. In Ex. XCIV the investigation introduces complex roots of equations in a way which commends them to common sense and makes them seem the natural complement of the real roots studied in Part I. In Ex. XCV it leads simply and convincingly to throwing open to directed numbers the one position from which they have hitherto remained excluded—namely, that of "base" in the logarithmic and exponential functions.

It is possible that the methods illustrated in these exercises are new; that is to say, the author discovered them independently and has not found them described elsewhere. It is, however, tempting to speculate that they are among those which Gauss had in his mind when he spoke of the possibility of giving to the whole arithmetic of complex numbers *die anschaulichsten Versinnlichung*. Whether this speculation is or is not well founded the teacher will, it is hoped, find that subjects which are usually rather dry and unprofitable gain greatly in vitality and intelligibility by the treatment here proposed for them.

§ 2. Ex. XCIV. *Complex Values of the Independent Variables.*—The argument of this exercise is given too fully to need additional explanation. In accordance with what was said above the student learns in division A that a quadratic equation which would have been declared in Part I to have no roots, or a cubic equation which would have been credited with only one root, may be conceived, after all, to have the full number of roots to which its degree entitles it —the supplementary roots being complex numbers that represent points not on the x-axis but on the plane through that axis perpendicular to the paper.

It is shown that the existence of complex roots as representable entities depends on the fact that the graph which exhibits the real roots is completed by another curve whose points lie outside the plane of the paper. The original curve and the supplement form together the tri-dimensional graph which represents the correspondence between the real values

of y and the complex values of x. *Exercises*, II, fig. 80, illustrates the character of this tri-dimensional graph when the function is parabolic. Fig. 113 below is the one to which reference is made in **No. 11.** The firm line is the circle on which lie all points which correspond to real values of both variables. The broken curve, turned so that its plane is perpendicular to the paper, is the hyperbola composed of the points which bring into correspondence real values of y and complex values of x.

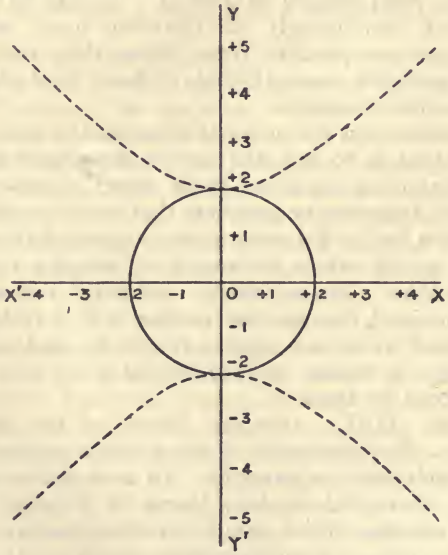


FIG. 113.

The adequacy of the representation is best demonstrated as follows. Since y is to be real only those values of x are admissible which make $4 - x^2$ positive or zero. If, therefore, x is real its value must not be above $+ 2$ or below $- 2$. Also since $y^2 + x^2 = 4$, all points which correspond to these real values of x lie on the circle of radius 2. On the other hand, if x is complex it must be of the form iv.[1] For this assumption makes $x^2 = - v^2$ and so makes $4 - x^2$ positive for all values of v. It will be seen that the assumption

[1] I.e. $0 + iv$, the "real" part of the complex being zero.

$x = u + iv$ would not work; for then x^2 would also be complex instead of being a real negative number. When $x = iv$ we have

$$y^2 = 4 + v^2$$
$$y^2 - v^2 = 4.$$

Since v means a length measured from the origin at right angles to the paper, and y, as usual, a length along the y-axis, it is evident that this relation describes a rectangular hyperbola lying in the plane through the y-axis perpendicular to the paper. The other examples of division B can be solved similarly.

§ 3. Ex. XCV. *Complex Values of a Function.*—In this exercise we consider cases in which real values of x are associated with complex values of y. The case taken as the basis of the investigation is of great interest and importance. In Section III the "growth-factor" r in the function $y = r^x$ was thought of as necessarily non-directed, or at least necessarily positive. The same fact came before us in Ex. LXXV, No. 11 (v), where the negative part of the graph of $y = x^x$ was found to degenerate into a few isolated dots on account of the impossibility of assigning values to a fractional power of a negative number (see fig. 95, p. 422). In Ex. XCV this limitation to the process of exponentiation is removed; it is found that fractional powers of a negative number do exist, but that they are complex. In other words, it is possible to have logarithms to a negative base, but their antilogarithms will, in general, be complex.

The argument leading to this conclusion is given fully in the exercise. Fig. 114 shows the diagram to be constructed in accordance with the directions in **Nos. 2, 3.** If a line were drawn smoothly through the points P, Q' and continued through the points marked on the circumferences of the other circles, the resulting curve would represent a plan or projection of the spiral $y = (-r)^x$ as viewed along its axis. The teacher is strongly recommended to supplement the drawings by the construction of a model of the spiral in the manner indicated in **No. 5.** The author's model was made in accordance with the following plan:—

A dozen empty cotton reels are first obtained, of identical shape and size and about an inch long; also a wooden rod which will just pass through the holes of the reels. Next, upon a sheet of fairly stout cardboard, thirteen circles are

drawn, the radius of the largest being 5 cms. and the radius of each of the others 0·9 of its predecessor in the series. Rectangles of diminishing size are then drawn round the circles. The largest measures 13·5 cms. by 10·4 cms., the smallest 10 cms. by 3·4 cms. Roughly the same proportions are observed in the intermediate rectangles. In each case the centre of the circle is 8·3 cms. from the lower edge of the card. (It is important that this dimension should be

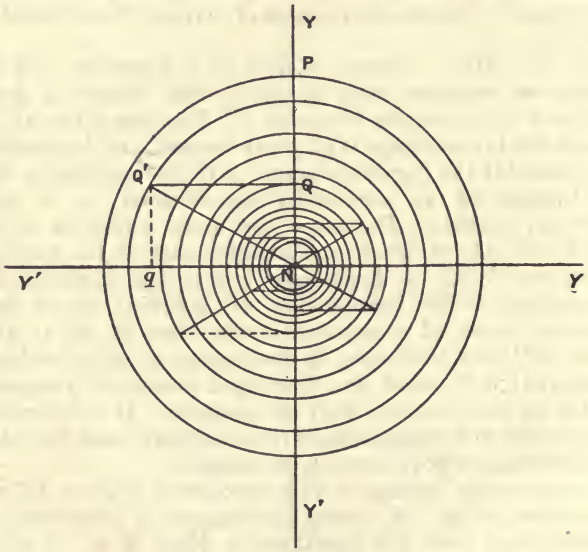


Fig. 114.

accurate.) Through the centre of each circle a line is drawn parallel to the longer sides of the rectangle. This line answers to YY′ in fig. 114. A mark is made on the circumference of the larger circle at the point corresponding to P in fig. 114, and is followed by others upon the other circles in succession in the positions indicated in fig. 114 and described in No. 3. The radii to the points thus marked need not be drawn, but it is obvious that they would make with the lines through the centres angles increasing regularly by 60°. The thirteen rectangles are now cut out. A hole

just large enough to admit a piece of firm but flexible wire[1] is made through each card at the point marked on the circumference of the circle. Another hole, to admit the wooden rod, is made near the bottom of each card. The distance of this hole from the bottom edge must be the same in each case, namely, a little greater than the radius of the cotton reels. The materials are now prepared for the construction of the model. One end of the rod is thrust through the hole at the bottom of the largest card and is forced tightly into a smaller hole in a cork or piece of wood held at the back of the card. A reel is now placed on the rod in contact with the face of the card. Thus the card now stands at right angles to the rod, held between the cork (or piece of wood) and the reel. The second card is next "threaded" on to the rod from the farther end, care being taken that its back is turned towards the face of the largest card. A second reel follows the second card down the rod so as to fix that card between two reels. The process is repeated with the successive reels and cards until the smallest card follows the twelfth reel. Reels and cards are now pressed tightly together and are fixed in position by a second closely fitting cork or block of wood. Lastly the wire is threaded carefully through the holes prepared for it, and the model is complete and exhibits two complete turns of the spiral. When viewed horizontally against a vertical background the spiral is seen projected as the curve $y = ar^{2x/\lambda} . \cos 2\pi x/\lambda$ (**No. 10**). When viewed from above it is seen projected as the curve $y = ar^{2x/\lambda} . \sin 2\pi x/\lambda$. In the author's model the spiral completes a turn in 18·4 cms. (i.e. $\lambda = 18{\cdot}4$), the radius of the largest circle is 5 cms. (i.e. a = value of y when x is zero $= 5$), and r is 0·9. Hence the formulæ of the projections are

$y = 5\ (0{\cdot}9)^{x/9{\cdot}2} . \cos \pi x/9{\cdot}2$ and $y = 5\ (0{\cdot}9)^{x/9{\cdot}2} . \sin \pi x/9{\cdot}2$. Expressed in the "standard form" (p. 310) the formulæ become $y = 5e^{-0{\cdot}0114} \cos \pi x/9{\cdot}2$ and $y = 5e^{-0{\cdot}0114} \sin \pi x/9{\cdot}2$.

The reader who is a physicist will know that these formulæ are of great importance in the theory of all vibratory phenomena.

The method of representation studied in division A can be

[1] The wire used by milliners in making the "shapes" of hats is much the best.

applied in all cases where consideration is confined to the complex values of y produced by real values of x. Division B contains several examples of this type in which the investigation is complementary to that of Ex. XCIV, Nos. 11-13. **No. 17** introduces a very interesting deduction ; namely, that a function whose real values are discontinuous for real values

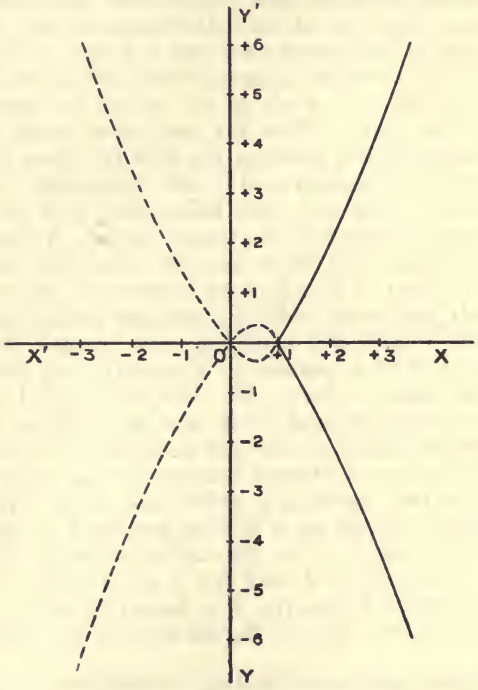


FIG. 115.

of x may yet be continuous if its complex values are taken into account. Fig. 115 shows the graph of Ex. LXXXV, No. 14, completed in this way. The broken line represents the " imaginary " values of y ; to get it into its proper position it must be twisted through a right angle about the x-axis.

If, in the argument of division A, r is made unity the spiral becomes one of constant radius and corresponds to the function $y = (-1)^x$. The very obvious properties of this

spiral are made, in division C, to assist investigation into that well-worn mathematical topic " the nth roots of ± 1 ". There is not much to be gained at this stage by an elaborate study of the subject, but Nos. 18-32 may be regarded as interesting " riders " to the general thesis of the exercise.

CHAPTER XLVI.

RELATIONS BETWEEN TWO COMPLEX VARIABLES.

§ 1. Ex. XCVI. *The Relations between two Complex Variables.*—We have seen (ch. XLIV., end of § 1) that complex numbers must be regarded as the typical numbers of algebra because they "form a universe complete in itself, such that, starting in it, we are never led out of it". In Ex. XCVI we begin a series of investigations which illustrate this important statement. The statement itself does not occur in the pages intended for the student; the teacher should, however, at an appropriate moment, direct attention to the facts which it summarizes.

In division A the student is asked in several instances to calculate values of y of the form $U + iV$, which correspond to values of x of the form $u + iv$. In division B he attacks the problem of representing graphically the correspondence between those values.

In **No. 9** we have

$$\begin{aligned} a^2 - x^2 &= a^2 - r^2 (\cos\theta + i \sin\theta)^2 \\ &= (a^2 - r^2 \cos 2\theta) - ir^2 \sin 2\theta. \end{aligned}$$

Assuming that the last expression is of the form

$$R(\cos\phi - i \sin\phi)$$

we have the equivalences

$$R \cos\phi = a^2 - r^2 \cos 2\theta, \qquad R \sin\phi = r^2 \sin 2\theta$$

whence

$$\begin{aligned} R &= \sqrt{\{(a^2 - r^2 \cos 2\theta)^2 + r^4 \sin^2 2\theta\}} \\ &= \sqrt{(a^4 + r^4 - 2a^2 r^2 \cos 2\theta)} \quad . \quad . \quad (1) \end{aligned}$$

and

$$\tan\phi = \frac{r^2 \sin 2\theta}{a^2 - r^2 \cos 2\theta} \quad . \quad . \quad . \quad . \quad (2)$$

In **No. 10** we start by assuming

$$a^2 - x^2 = R(\cos\phi - i \sin\phi)$$

where R and ϕ have the values just determined. It follows that

$$\begin{aligned} y &= \sqrt{(a^2 - x^2)} \\ &= \sqrt{\mathrm{R}} . (\cos \phi - i \sin \phi)^{\frac{1}{2}} \\ &= \sqrt{\mathrm{R}} . (\cos \tfrac{1}{2}\phi - i \sin \tfrac{1}{2}\phi) \quad . \quad . \quad (3) \end{aligned}$$

Given the values of r and θ this expression can always be turned into a numerical form by means of relations (1) and (2). If y is to be real sin $\frac{1}{2}\phi$ must be zero. That is, $\frac{1}{2}\phi = s\pi$, where s is any whole number, and $\phi = 2s\pi$. It follows that tan $\phi = 0$, and, therefore, from (2) above, that sin $2\theta = 0$. Hence $2\theta = s\pi$, where s is any whole number, and $\theta = s\pi/2$. Now if $2\theta = s\pi$, cos 2θ must be either + 1 or − 1 according to whether s is odd or even. Hence when y is real we have from (1)

$$\begin{aligned} \mathrm{R} &= \sqrt{(a^4 + r^4 \pm 2a^2r^2)} \\ &= (a^2 \pm r^2) \end{aligned}$$

no alternative signs being required before the bracket since R is a "modulus" and therefore non-directed. Again, since $\frac{1}{2}\phi = s\pi$, cos $\frac{1}{2}\phi = \pm 1$. So relation (3) finally takes the form

$$y = \pm \sqrt{(a^2 \pm r^2)} \quad . \quad . \quad . \quad (4)$$

Comparing this result with the formula

$$y = \sqrt{(a^2 - x^2)}$$

we see that $x = \pm r$ when the sign before the r^2 in (4) is *minus*, and $x = \pm ir$ when the sign is *plus*. This conclusion was obtained by much simpler considerations in Ex. XCIV, No. 11 (see p. 480).

The principles underlying the graphic representation of the relations between two complex variables are fully explained in § 2 of the exercise. The method is due ultimately to Riemann.[1]

Two details should be noted. (i) Since all the points of each of the representative planes are involved in the correspondence some special plan must be adopted of presenting the scheme of relations to the eye. The principle already applied in map projections and in the more general "transformations" of Ex. XCI, B, is the one to which it is natural to have recourse. (ii) It is convenient to have a special notation to show when complex values of the variables are under con-

[1] *Grundlagen fiir eine allgemeine Theorie der Functionen einer verändlichen complexen Grosse* (1851).

sideration. (Strictly speaking, complex values should be regarded as normal and restriction to real values regarded as the exception; but in an elementary work it is more useful to consider that the variables are real unless the contrary is indicated.) For this purpose the notation $\dot{x}$, $\dot{y}$ has been adopted. This departure from the usual symbolism [1] is adopted in order to emphasize the continuity of ideas between the cases in which the variables are real and those in which they are complex. The diacritic dot would more conveniently be placed above the letter, but the symbols would then be identical with Newton's notation for "fluxions," which is still employed

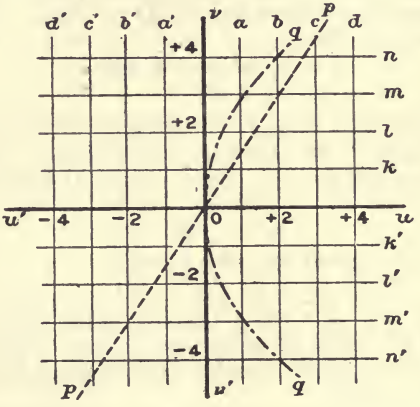


FIG. 116.

by many writers and is likely to become more rather than less usual.

Fig. 116 is the $\dot{x}$-net of **Nos. 11-13**; fig. 117 the net into which the u-lines and v-lines of fig. 116 are transformed in accordance with the function $\dot{y} = \frac{1}{4}\dot{x}^2$. It will be observed that corresponding positive and negative v-lines (e.g. those lettered a and a') transform into a single parabola—or rather into two coincident parabolas of which one "begins," so to speak, above and the other below the U-axis. A similar observation applies to the u-lines. In fig. 116 the broken line

[1] Riemann used $z = x + iy$ for the independent variable and $w = u + iv$ for the dependent variable.

pOp is $v = 1{\cdot}5u$. It transforms in fig. 117 into a straight line doubled upon itself at O. The parabola in fig. 116 passing through the points where $u = \frac{1}{3}v^2$ transforms in fig. 117 into the cusped curve qOq'.

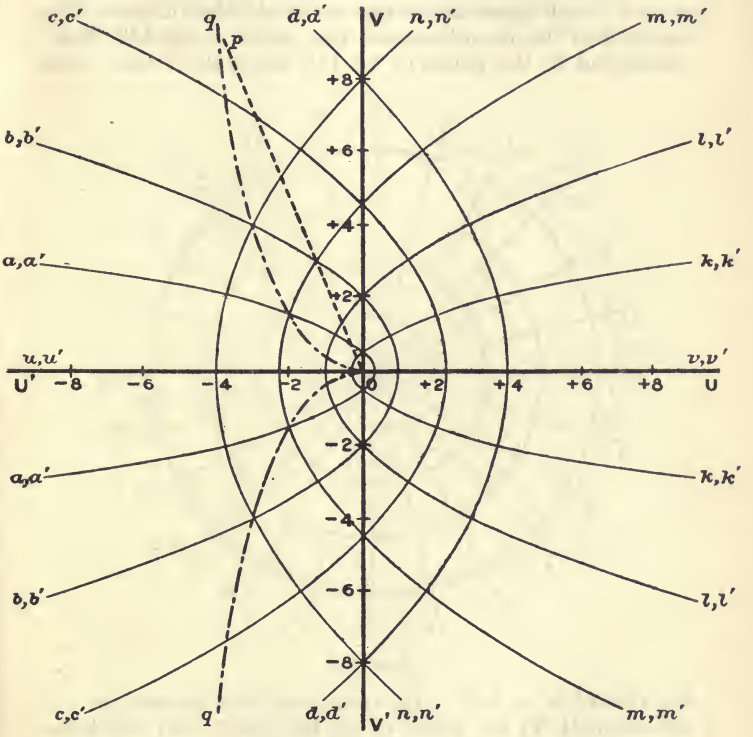


FIG. 117.

The method of obtaining the graphic transformation is as follows. We have

$$\begin{aligned} y &= \tfrac{1}{4}x^2 \\ &= \tfrac{1}{4}(u + iv)^2 \\ &= (u^2 - v^2)/4 + iuv/2 \end{aligned}$$

whence $\quad U = (u^2 - v^2)/4$ and $V = iuv/2$.

Let u have the constant value + 1 and substitute for v in succession 0, + 1, − 1, + 2, − 2, . . . In this way the values are obtained of the co-ordinates (u, v) of the points in fig. 117 which correspond to the points in fig. 116 where the horizontal v-lines cross the vertical $u = + 1$. Next put $u = + 2$ and repeat the former series of substitutions. The results give the co-ordinates of the points in fig. 117 which correspond to the points in fig. 116 where the v-lines cross

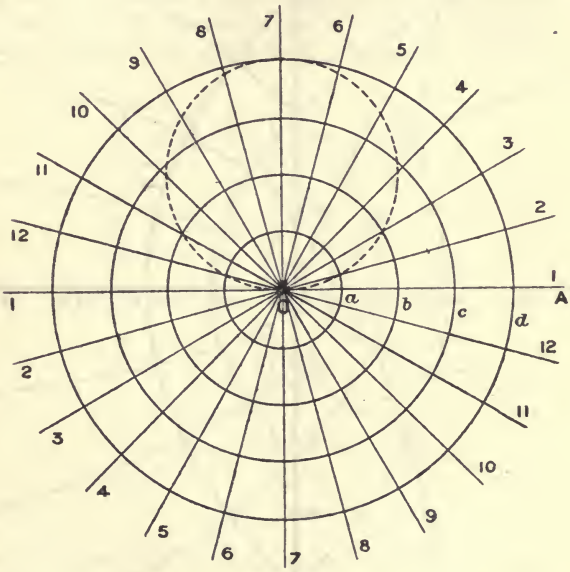


FIG. 118.

the vertical $u = + 2$. By continuing this process the co-ordinates (U, V) are found of all the points into which the crossing points in fig. 116 are transformed. When these are inserted in their proper places on the graph-paper the two sets of parabolas which represent the original u-lines and v-lines can be drawn at once. The transformations of the lines $p\mathrm{O}p$ and $q\mathrm{O}q$ in fig. 116 are obtained by marking in fig. 117 the points which correspond to those in which the lines cross the u-lines and v-lines. For this purpose the lines are treated exactly like lines of latitude and longitude in map-

drawing. When the points have been inserted in fig. 117 the graphs are drawn through them in the usual way.

Figs. 118 and 119 give the solutions to **Nos. 14-16.** Fig. 118 shows the netting of the x-plane, the θ-lines being numbered from 1 to 12 and the r-circles lettered from a to d for ease of identification in the transformed figure. In this case we have

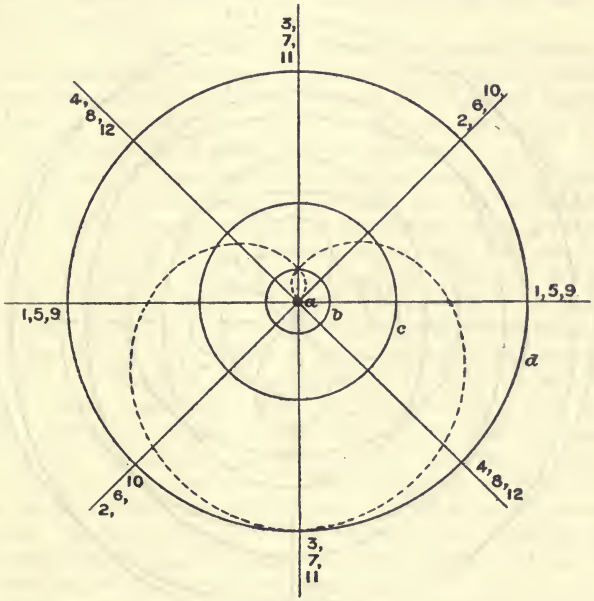


Fig. 119.

$$y = 0{\cdot}1x^3$$
$$= 0{\cdot}1r^3(\cos\theta + i\sin\theta)^3$$
$$= 0{\cdot}1r^3(\cos 3\theta + i\sin 3\theta).$$

Thus in the net of the y-plane the line corresponding in fig. 118 to a certain value of θ is transformed into a line making with the initial line an angle 3θ, while a given r-circle is transformed into a circle in which $R = r^3/10$. In fig. 118 the points for which the relation $r = 4\sin\theta$ holds good lie upon the broken circle. To find the transformation—the

beautiful broken curve of fig. 119—the points are marked in the second figure which correspond to the points in the first where the circle crosses the θ-lines and r-circles.

In **No. 16** r is transformed into $\sqrt{R}$ and ϕ into $\phi/2$, the symbols having the meanings assigned to them in No. 9. By means of these relations the transformed co-ordinates must be found for each point where the r-circles and θ-lines of

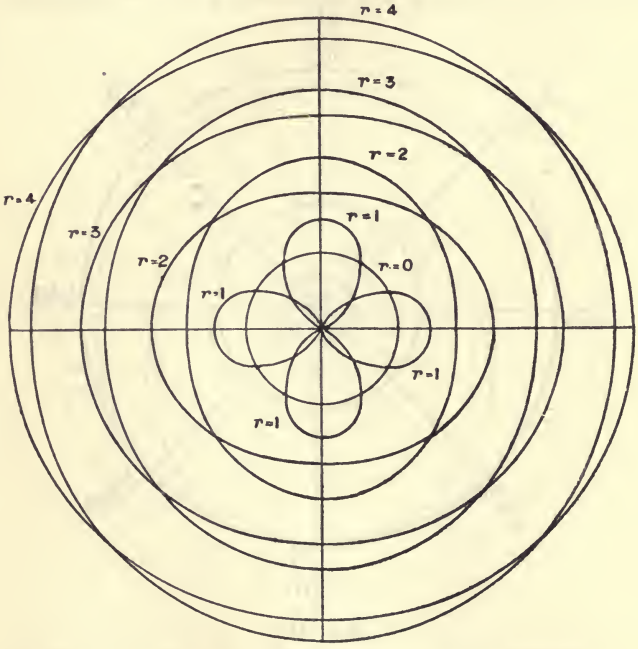


FIG. 120.

fig. 118 intersect. The calculation is a useful one if it can be shared among several computers; otherwise it is too laborious to be worth undertaking. As a specimen of the work take the transformation of the point where the circle of radius 2 crosses the line $\theta = 30°$. To determine $\phi/2$ we substitute $\theta = 30°$, $r = 2$ in formula (2) on p. 486. The result is a double one; the value of $\phi/2$ is either 53° 3′ or 143° 3′. To find $\sqrt{R}$ we substitute $\theta = 30°$ and $r = 2$ in formula (1)

on p. 486 and obtain $\sqrt{R} = 1{\cdot}9$. Thus two points, lying on lines at right angles must be plotted in the y-plane to represent this particular point in the x-plane. Similar results hold good for the other intersections. The consequence is that each of the r-circles transforms into two curves (fig. 120) of which one is the other rotated through a right angle.

Fig. 120 shows the transformations of the r-circles, the transformations of the θ-lines being omitted for the sake of greater clearness.[1]

§ 2. Ex. XCVII. *The Logarithm of a Complex Number.*— When a function is expressible as a sum of a definite number of powers of x or a root of such a sum De Moivre's theorem makes it evident that to every value of x of the form $u + iv$ there must correspond a value of y of the form $U + iV$. This fact has, perhaps, been sufficiently illustrated in the previous exercises. The same thing can be seen, but not so easily, to be true of more complicated functions expressed in terms of a definite number of powers of x, root signs, etc. But when we come to functions, such as the logarithm, the sine, etc., in which y is not expressed in terms of powers of x, the matter is very different; it is not by any means evident that for every complex value of x there exists a corresponding complex value of y. The next two exercises are intended in the first place to show that what is true of functions of the "rational" and "algebraic" types is also true of the typical "transcendental" functions. In this way the evidence in favour of Cayley's generalization (p. 474) is greatly widened. In the second place the exercises are intended to emphasize certain incidental results of the analysis—for example, the "exponential values" of the sine and cosine—which are themselves of considerable importance.

The question of the logarithmic and exponential functions of a complex variable is not an easy one, and it is probable that the line of argument followed in Ex. XCVII may not satisfy the more exacting mathematician. It is believed, however, that it is not actually faulty from the logical stand-

[1] By an error, detected too late for correction, the transformation of $r = 0$ is represented in fig. 120 as a circle of unit radius. When $r=0$, $\tan \phi = 0$ by (2), p. 486; hence $\phi/2 = s\pi/2$. Also from (1) $\sqrt{R} = 1$. Hence the origin in fig. 119 transforms into four points situated where the circle drawn by error in fig. 120 cuts the axes.

point, and that it will prove easy and interesting. The teacher who seeks a more rigorous and complete treatment should consult Mr. Hardy's *Pure Mathematics*, ch. x.

Nothing need be added to the full exposition of the exercise except the remark that the assumption, $k = 1$, which is professedly made in § 2 for the sake of convenience, is justified in Ex. XCVIII, Nos. 7, 8.

In **No. 9** if we put $\pi x = \theta$ we have $x = \theta/\pi$ and

$$\cos\theta + i\sin\theta = (-1)^{\theta/\pi} = i^{2\theta/\pi}$$

whence $$\log(\cos\theta + i\sin\theta) = \frac{2\theta}{\pi}\log i.$$

But if $\log(\cos\theta + i\sin\theta) = i\theta$ also, then we have

$$\frac{2\theta}{\pi}\log i = i\theta$$

or $$\log i = i\frac{\pi}{2}.$$

In **No. 10** we have, by our definition of the logarithm of a complex number, that

$$\log ib = \log b\,(\cos\pi/2 + i\sin\pi/2) = \log b + i\frac{\pi}{2} = \log b + \log i$$

by the result of No. 9. Since this identity in the results could follow only from the assumption that

$$\log(\cos\theta + i\sin\theta) = ik\theta$$

it may be taken as a strong confirmation of that assumption and therefore as a partial confirmation of the validity of the assumption

$$\log(\cos\theta + i\sin\theta) = i\theta.$$

§ 3. Ex. XCVIII. *Supplementary Examples.*—The short supplementary exercise is devoted to two related topics, the "exponential values" of the sine and cosine and the question of finding values for the sine and cosine of a complex variable. It is to be noted that the exercise is not intended to be taken until the student has learnt the exponential expressions for the hyperbolic functions. The corresponding expressions for the circular functions are then welcomed for the sake of their analogy with the former ones. Once more

the teacher is warned that a number of important subtleties are ignored; for these he is again referred to works such as Mr. Hardy's *Pure Mathematics* or Chrystal's *Algebra*, Part II.

In **No. 7** the differential formulæ are (by Ex. LXXXIII)

$$\text{(i)}\quad \frac{\delta y}{\delta x} = -\frac{1}{2i}(e^{ix} - e^{-ix})$$

and

$$\text{(ii)}\quad \frac{\delta y}{\delta x} = \tfrac{1}{2}(e^{ix} + e^{-ix}).$$

For *if* i *may be treated as if it were a number*, e^i may be replaced by a single symbol a, and we have $e^{ix} = a^x$. Similarly e^{-ix} may be written b^x where $b = e^{-i}$. Thus we have in the first of the two given cases

$$y = \tfrac{1}{2}(a^x + b^x)$$

and

$$\frac{\delta y}{\delta x} = \tfrac{1}{2}(a^x . \log a + b^x . \log b)$$

$$= \tfrac{1}{2} i (e^{ix} - e^{-ix})$$

$$= -\frac{1}{2i}(e^{ix} - e^{-ix}).$$

But the formula of No. 7 (i) is the exponential value of cos x and the differential formula derived from it the exponential expression for $-\sin x$. Hence differentiation of the exponential value produces the result which theory teaches us to expect. The same thing is true of the other formula. On the other hand, when we seek the differential formula of **No. 8** we find that it has the form

$$\frac{\delta y}{\delta x} = \tfrac{1}{2}ik(e^{ikx} - e^{-ikx}).$$

Now the interest of this result is that it shows that the formula of No. 8 cannot be a valid expression for cos x, for the corresponding differential formula is not the exponential expression for $-\sin x$. Thus we are entitled to conclude that the assumption $k = 1$ made in *Exercises*, II, p. 208, was not only permissible but actually necessary. No other assumption would have guaranteed us against inconsistency in our arguments.

Fig. 121 exhibits the transformations of the u-lines and v-lines of fig. 116 in accordance with the function (**No. 14**)

$$y = \sin x.$$

The figure was constructed as in the former cases by calculating (by the equivalence proved in No. 13) the values of U and V (i.e. $\sin u \cdot \cosh v$ and $\cos u \cdot \sinh v$) for each of the

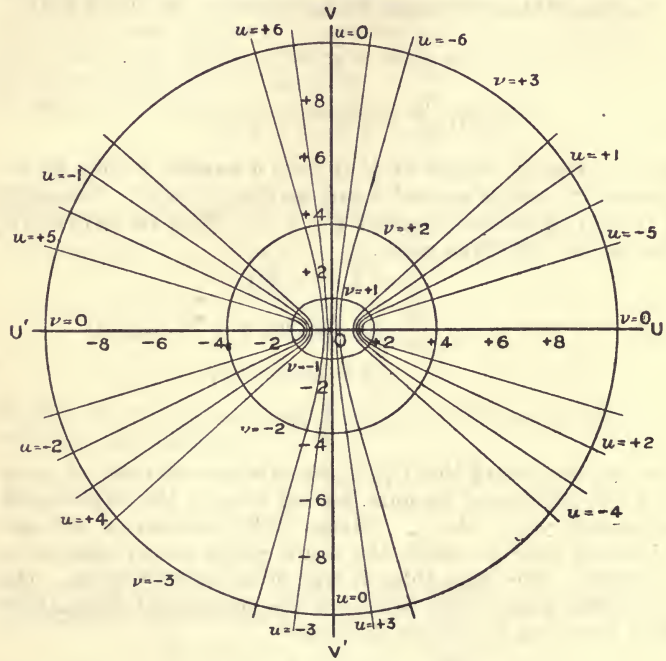


FIG. 121.

intersections of the u-lines and v-lines in fig. 116 and drawing smooth curves through the points thus obtained. The various lines can be identified by the numbers attached to them and should be carefully studied for the sake of the remarkable results which they bring out.

SECTION VII.

PERIODIC FUNCTIONS.

THE EXERCISES OF SECTION VII.

*** The numbers in ordinary type refer to the pages of *Exercises in Algebra*, Part II ; those in heavy type to the pages of the present volume.

EXERCISE		PAGES
XCIX.	CIRCULAR MEASURE	217, 499
C.	ANGLES OF UNLIMITED MAGNITUDE	224, 501
CI.	SUM AND DIFFERENCE FORMULÆ	235, 508
CII.	CIRCULAR FUNCTIONS	243, 512
CIII.	INVERSE CIRCULAR FUNCTIONS	248, 514
CIV.	PROGRESSIVE WAVE MOTION	254, 515
CV.	STATIONARY WAVE MOTION	264, 518
CVI.	HARMONIC ANALYSIS	273, 520
CVII.	DIFFERENTIAL FORMULÆ FOR PERIODIC FUNCTIONS	285, 528
CVIII.	HYPERBOLIC SINES AND COSINES	293, 531
CIX.	HYPERBOLIC FUNCTIONS	301, 534
CX.	SUPPLEMENTARY EXAMPLES—	
	A. Differential formulæ and expansions of the sine and cosine	306, 528
	B. Differential formulæ for $\tan x$	310, 529
	C. The calculation of π	311, 530
	D. The Gudermannian functions	313, 536
	E. The prediction of Tides	315, 526

CHAPTER XLVII.

THE CIRCULAR FUNCTIONS.

§ 1. Ex. XCIX. *Circular Measure.*—The reason (and the only reason) for measuring angles in radians rather than degrees is precisely the same as the reason for preferring logarithms to base e to common logarithms—namely, that theoretical arguments become much simpler. For example, the "differential coefficient" of $\sin\theta$ is $\cos\theta$ if θ is the number of radians in an angle while it is $\frac{\pi}{180}\cos\theta$ if θ is a number of degrees. On this account the study of circular measure has been deferred until the student is ready to make use of this superiority, and is introduced by a method which brings out its advantages at the outset. The essence of the argument is to show (i) that when an angle is small its circular measure may be substituted for its sine (or tangent) in approximate calculations, and (ii) that when it is not small the values of the sine and cosine may still be expressed approximately in terms of its circular measure by means of the simple formulæ

$$\sin\theta = \theta - \frac{\theta^3}{6}$$

and

$$\cos\theta = 1 - \frac{\theta^2}{2}\left[+ \frac{\theta^4}{24}\right].$$

This interesting result is reached by a simple graphic method. The method does not *prove* the equivalence

$$\sin\theta = \theta - \frac{\theta^3}{6} +$$

but it *suggests* that formula as a simple way of expressing approximately the value of the sine of an angle in which $\theta \not> 1$. Thus it prepares the student for the demonstration

given at a later stage (Ex. CX, A). Meanwhile it enables him to apply the notion of an expansion of the sine (or cosine) to the solution of various problems, and so to realize the great advantages of radian measure.

The teacher will note the proposal (Ex. XCIX, § 1) to confine the symbols θ and ϕ to circular measure and to use α, β and γ as symbols for measurements in degrees. The distinction is a useful one and should be maintained systematically.

Division A of the exercise offers no difficulties. The argument of division B, Nos. 18-20, runs as follows :—

No. 18. $\cos\theta = \sqrt{1 - \sin^2\theta}$

$$= \sqrt{\{1 - \left(\theta - \frac{\theta^3}{6}\right)^2\}} \text{ approx.}$$

$$= \sqrt{\{1 - \theta^2\left(1 - \frac{\theta^2}{3} + \frac{\theta^4}{36}\right)\}} \text{ approx.}$$

$$= 1 - \frac{\theta^2}{2}\left(1 - \frac{\theta^2}{3} + \ldots\right) + \qquad \text{[Pt. I, pp. 72, 73.]}$$

$$= 1 - \frac{\theta^2}{2} + \ldots$$

Since subsequent terms will all involve θ^4 or higher powers they may be neglected.

Nos. 19, 20. By No. 18 $\cos^2\theta = 1 - \theta^2 + \frac{\theta^4}{4} +$

and $\sin^2\theta = \theta^2 - \frac{\theta^4}{3} +$

$$\therefore \sin^2\theta + \cos^2\theta = 1 - \frac{\theta^4}{12} +$$

Now we know that the equivalence for $\sin^2\theta$ is correct as far as it goes. It follows that the equivalence $\cos^2\theta = 1 - \theta^2 + \theta^4/4$ is not adequate as for the term involving θ^4. To make it so we must assume that the expression for $\cos\theta$ contains a term such that, when the expression is squared, the coefficient of θ^4 becomes $\frac{1}{3}$ instead of $\frac{1}{4}$. Assume, therefore,

$$\cos\theta = 1 - \frac{\theta^2}{2} + \frac{\theta^4}{a} +$$

Then $$\cos^2\theta = 1 - \theta^2 + \left(\frac{2}{a} + \frac{1}{4}\right)\theta^4 + \ldots$$

and we must have $$\frac{2}{a} + \frac{1}{4} = \frac{1}{3}$$

whence $$a = 24.$$

We conclude that the completer approximation required is

$$\cos\theta = 1 - \frac{\theta^2}{2} + \frac{\theta^4}{24}.$$

§ 2. Ex. C. *Angles of Unlimited Magnitude.*—The purpose of Ex. C is to introduce and illustrate the notions (i) that an "angle" may be of any magnitude, positive or negative, (ii) that to every angle there correspond a sine, a cosine, a tangent, etc., and (iii) that the values of the ratios connected with the angle are repeated endlessly in cycles as the value of the angle rises or sinks. In divisions A and B these notions are introduced and applied to the analysis of simple vibratory motion, such as the swinging of a pendulum. In divisions C and D their wider geometrical applications are illustrated. Thus the exercise is one of critical importance.

The reason for basing the extension of the angle-concept upon a study of harmonic motion has already been given. It is not necessary actually to carry out the experiment of Ex. C, § 1, though it is well (and easy) to do so. Many school laboratories possess, in the shape of a flat steel spring and a Fletcher trolley, apparatus specially adapted for performing it luxuriously. Much simpler contrivances will, however, suffice. The text describes one which may be simplified further by the omission of the vice as follows: Take a lath long and thin enough to vibrate widely and fairly slowly. (Nothing serves better than one of the flat iron laths used to support beds before the era of the spring mattress. They can generally be obtained at a second-hand furniture shop.) Fasten the paint brush firmly across it at one end; place the other end upon a flat box or pile of books a few inches high, resting on the teacher's desk; add another box, block or book and let a boy or girl press steadily upon it to keep that end of the lath at rest. Pull the other end two or

three inches vertically up or down and release it. Hold the drawing board with its plane vertical so that the tip of the brush just plays over the surface of the paper. Move the board steadily along the desk. In this way several "periods" of the curve of *Exercises*, II, fig. 87, can be obtained with little loss of amplitude.

Two other methods may be indicated, partly on the ground of their simplicity and partly because they bring out the fact that "pendulum motion" is identical with harmonic motion when the pendulum swings through a sufficiently small angle. One of these is described in Ex. CI, § 2, p. 239, and its use at this point would be preparatory for the further use to be made of it when that exercise is reached. The other is as follows: Pin a sheet of paper to a drawing board and fasten two strings of equal length at the corners on one of the longer sides of the board. Fasten the free ends of the strings to a gas bracket, a map holder or in any other way so that the board may hang and move freely. At worst a boy or girl, standing on a chair, may hold them. Draw the board carefully aside in its own plane. At the same time let some one hold a brush (dipped in red ink) horizontally so that the tip touches the paper. Now release the board and, as it swings, let the brush be steadily raised or lowered with the tip always just touching the paper. A few trials will give sufficient expertness in deciding the proper length of swing and rate of movement of the brush.

The harmonic curves obtained by any of these methods should be measured so as to show that their ordinates follow, at any rate roughly, the law

$$b = a \sin \frac{360^\circ}{l} d.$$

It will be obvious that deviations from this law will necessarily be caused (i) by the uncertainty introduced by the thickness of the line, (ii) by uneven movement of the board or the brush, and (iii) by the inevitable decay in the amplitude of the swings in all cases. On the other hand it will be understood that *Exercises*, II, fig. 87, is the ideal curve which would be obtained if these disturbing circumstances could be eliminated.

The foregoing experiments supply in a clear form the data needed for our investigation. The "wheel" imagined in § 2 of the exercise is a device for *analysing* these data. The

teacher who distrusts his executive powers may (as we have said) omit the former experiments altogether. The experiment of § 2 is more important because it illuminates so clearly the idea of a continuous increase of the angle associated with periodic recurrences of the cycle of sines. The simplest piece of apparatus will suffice. Take a cork, bore it with a hole passing through the centres of its flat ends. Thrust a lead pencil through the hole to act as the axle of the wheel and fix a hat-pin (4 to 6 inches long) into the cork so that it is perpendicular to the pencil. Hold it above the head near a wall (or in front of a sheet of cardboard fixed vertically upon the desk) so that the pencil is horizontal and parallel with the wall (or card). Twist the pencil uniformly between the fingers, and the shadow of the head, cast by a distant light in the proper position, will exhibit the required S.H.M.

The argument of §§ 1, 2 of the exercise and examples Nos. 1-10 are of extreme importance but require no further comment. In **No. 11**, the extended definition of the circular measure is, of course, as follows: Let a line OP of unit length, starting from the usual initial position along OX, rotate about O any number of times either in the positive or the negative direction, and finally come to rest in any position. Then it shall be considered to have traced out an angle whose circular measure, θ, is the distance through which the point P has moved. If the direction of rotation was anti-clockwise θ is reckoned positive, if clockwise, negative.

It follows from this definition that there is no single angle corresponding to a single position of the line OP. Every position is associated with an endless series of angles, positive and negative, consecutive members of which differ constantly by 2π. If we want to remove this ambiguity we must use some such device as that of the German mathematician Riemann (*c.* 1857). When OP has travelled completely round the circle in the course of any one revolution, it may be supposed to move during its next revolution, not over the same surface as before, but over another surface which is pressed down so that it is, to all intents, in the same place as the former. Thus OP moves over an endless spiral surface arranged around the perpendicular to the paper through O like the turns of a spiral staircase, but compressed from above and below so that all the endless series of spires appears in the plane of the paper. If we adopt this artificial idea we may say

that a single angle corresponds to every single position of OP, for as OP revolves it will never pass for a second time over the same surface.

The teacher is left to decide whether he will add this idea to those given in the text.

§ 3. Ex. C, C. *Spirals.*—The practical value of the idea that an angle may have any magnitude is further illustrated in divisions C and D. In fig. 122 the firm line is an Archimedes' spiral (**No. 21**) in which a and θ are positive, the dotted line one in which a and θ are negative. Fig. 123 shows the first positive and the first negative turns of the logarithmic spiral of **No. 22**, and parts of the second turns. The tracing point

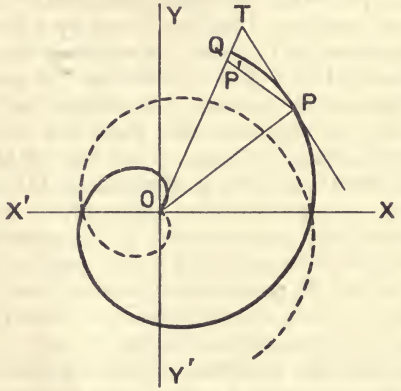


FIG. 122.

starts from A, where $OA = a^0 = 1$. In accordance with ch. xxx., p. 317, a must always be positive. The spiral of **No. 23** will be identical with fig. 123 but inverted. Since $r = a^\theta$ is merely another way of writing $\log_a r = \theta$ it is obvious that, by means of a centimetre rule and a protractor, either spiral can be used to give logarithms to base a.

The lines OP, OP′, etc., added to fig. 122 correspond to the argument of § 3. To answer **No. 24** we note that the angle OPT ($= \phi$) at which the curve cuts the radius vector is the sum of $\angle OPP'$ and $\angle TPP'$. As Q comes nearer to P $\angle OPP'$ and $\angle OP'P$ approach, and can be made as little different from a right angle as we please. Hence we have, to as close a degree of approximation as we please,

$$\phi = \frac{\pi}{2} + \angle \mathrm{TPP}'$$

and therefore $$\cot \phi = -\tan \mathrm{TPP}'$$

$$= -\frac{1}{r} \cdot \frac{\delta r}{\delta \theta}.$$

In **No. 25** we have $r = a\theta$, $\delta r/\delta\theta = a$. Hence

$$\cot \phi = -\frac{1}{r} \cdot \frac{\delta r}{\delta \theta}$$

$$= -\frac{1}{\theta}.$$

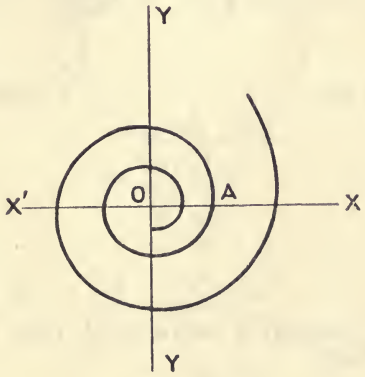


FIG. 123.

In **No. 26** $\delta r/\delta\theta = a^\theta \cdot \log_e a = r \cdot \log_e a$ [Ex. LXXXIII, No. 5], so that

$$\cot \phi = -\log_e a.$$

To prove the area-formula of **No. 27** we note that (fig. 122)

$$\text{area of OPQ} = \tfrac{1}{2}\mathrm{OQ} \times \mathrm{PP}'$$

more nearly the more nearly PP′ is perpendicular to OQ. When $\angle$POQ is so small that we may call it $\delta\theta$ we can at the same time write

$$\mathrm{OQ} = r + \delta\theta$$

and $$\text{area of OPQ} = \tfrac{1}{2}(r + \delta\theta) \cdot r\delta\theta$$

$$= \tfrac{1}{2}r^2 \cdot \delta\theta$$

since, by the definition of a differential, $(\delta\theta)^2$ may be omitted.

The hyperbolic spiral of **No. 31** is shown in fig. 124. As θ approaches zero the curve approaches the line A distant a

from the x-axis. As in Archimedes' spiral, a and θ must have the same sign in order that r may always be positive. The area swept out by the radius-vector between two positions θ_1 and θ_2 is found from the differential formula

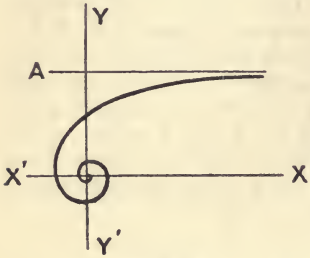


FIG. 124.

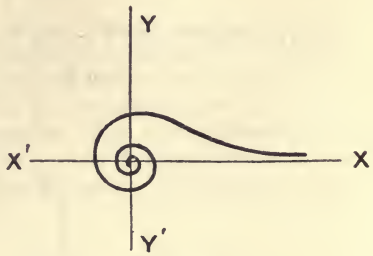


FIG. 125.

$$\delta A = \tfrac{1}{2} r^2 \,.\, \delta\theta$$
$$= \frac{1}{2} \cdot \left(\frac{a}{\theta}\right)^2 \cdot \delta\theta$$
$$= \frac{a^2}{2} \cdot \frac{\delta\theta}{\theta^2}$$

whence $$A = \frac{a^2}{2}\left(\frac{1}{\theta_1} - \frac{1}{\theta_2}\right)$$ by Wallis's Law.

The lituus (**No. 32**) is shown in fig. 125. For the area-problem we have

$$\delta A = \tfrac{1}{2} r^2 \,.\, \delta\theta$$
$$= \frac{1}{2} \frac{a^2}{\theta} \,.\, \delta\theta$$

whence $$A = \frac{a^2}{2} \log (\theta_2/\theta_1)$$ [Ex. LXXXIII, No. 26.]

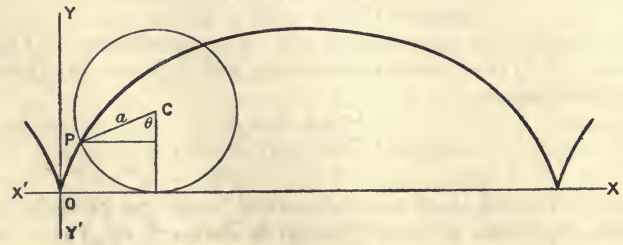


FIG. 126.

§ 4. Ex. C, D. *Roulettes.*—Fig. 126 exhibits one complete period of the cycloid of **No. 34** together with the construction

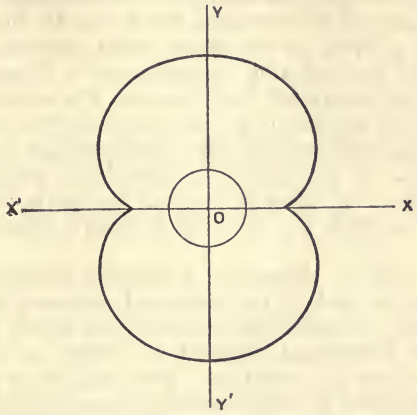


Fig. 127.

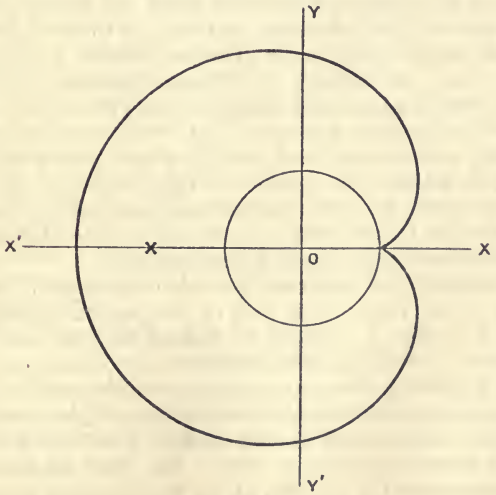


Fig. 128.

necessary to prove the formulæ of **No. 33.** Figs. 127, 128 show the epicycloids of **Nos. 37** and **38.**

It is easy to verify the theory of the cycloid by the simple method explained in Exercises Pt. I, Ex. II, No. 30. The epicycloids of **Nos. 37** and **38** are often seen as "caustics" produced by the reflexion of light from a circular mirror. Thus half the epicycloid in which $n = 2$ is seen on the surface of milk in a glass, or on the tablecloth within a polished serviette ring whenever the illumination comes from a single distant light. If the light (e.g. a match) is brought close up to the edge of the glass or the ring the curve becomes the epicycloid in which $n = 1$. (See Preston's *Theory of Light*, ch. v.)

§ 5. Ex. CI, A. *Sum and Difference Formulæ.*—Before free use can be made of the sines and cosines of angles $> 2\pi$ it is necessary to determine whether they follow the laws of combination established in Part I, Ex. LXII, for angles $< 2\pi$ whose sum does not exceed 2π. This question is investigated in Ex. CI, division A. The method adopted depends upon the facts (i) that when we know the sine and cosine of an angle θ' in the first quadrant we can at once write down the sines and cosines of the angles in the second, third, and fourth quadrants which are obtained by adding one, two, or three right angles to θ; (ii) that the addition of a further number of right angles simply brings OP into one of its former positions. Now let θ' and ϕ' be two angles each less than $\pi/2$ and therefore subject to the laws of combination of Ex. LXII. Obtain a new angle, θ, by adding any number of right angles to θ', and another new angle ϕ by adding any number of right angles to ϕ'. Then by means of the above-mentioned relations between $\sin\theta$ and $\sin\theta'$, $\cos\theta$ and $\cos\theta'$, etc., it is possible to show that the laws of combination apply to every possible case.

To establish this statement different members of the class are first set to verify it for each of the possible sixteen cases indicated in **No. 3.** (There are sixteen because each of the four values of θ can be taken with any one of the four values of ϕ.) In writing out in an examination the proof of any one of the equivalences it would, of course, be sufficient to indicate the field of cases, to prove one, and to state that the others could be proved in the same way. The demonstration must then be completed, as in **No. 11,** by showing that the equiva-

lence also holds good when θ and ϕ are negative—singly or together.

In **No. 8** $(\theta + \phi)$ is of the form $2n\pi + (\theta' + \phi')$ where $2n$ means simply "some even number". Hence

$$\begin{aligned}\cos(\theta + \phi) &= \cos(\theta' + \phi') \\ &= \cos\theta' \cos\phi' - \sin\theta' \sin\phi' \\ &= \cos\theta \cos\phi - \sin\theta \sin\phi.\end{aligned}$$

For since $\theta = 2n\pi + \theta'$, $\cos\theta = \cos\theta'$, $\sin\theta = \sin\theta'$, etc.

In **No. 10** consider the equivalence for $\sin(\theta - \phi)$ when $\theta = (2n + \frac{1}{2})\pi + \theta'$ and $\phi = (2n + 1)\pi + \phi'$. In this case

$$\sin\theta = \cos\theta', \cos\theta = -\sin\theta',$$
$$\sin\phi = -\sin\phi', \cos\phi = -\cos\phi'.$$

Also $\theta - \phi$ is of the form $(2n + \frac{3}{2})\pi + (\theta' - \phi')$,[1] so that

$$\begin{aligned}\sin(\theta - \phi) &= -\cos(\theta' - \phi') \\ &= -\cos\theta' \cos\phi' - \sin\theta' \sin\phi \\ &= (-\sin\theta).(-\cos\phi) - (-\cos\theta).(-\sin\phi) \\ &= \sin\theta \cos\phi - \cos\theta \sin\phi.\end{aligned}$$

The division ends with a few identities most of which are frequently needed in future work.

§ 6. Ex. CI, B. *Compound Harmonic Motion.*—This subject has not the same fundamental importance as simple harmonic motion and may be omitted without serious loss. It is included as offering interesting application for formulæ in which the angles may have any magnitude.

The curves produced by compounding rectangular harmonic motions are called "Lissajou's curves" and are of some importance in the theory of sound. The "compound pendulum" produces many of them in a simple and sufficiently effective way but cannot be used to compound S.H. motions of the same frequency. The teacher who is in touch with a physical laboratory should be able to obtain assistance in illustrating this topic in a more elaborate way if he feels disposed to do so.

The solution of **No. 22** is, of course,

$$y/x = (b \sin\theta)/(a \sin\theta) = b/a.$$

That is, the resultant vibration is in a straight line making with the x-axis the angle whose tangent is b/a. In **No. 23** the line lies symmetrically on the other side of the y-axis. In **No. 24** we have

[1] Remember that $2n$ in the expressions for θ and ϕ and $\theta + \phi$ means "even number" but not necessarily the same even number.

$$x^2/a^2 = \sin^2\theta,\ y^2/b^2 = \cos^2\theta$$
$$\therefore \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1.$$

In **No. 26** we have

$$\sin\theta = x/a,\ \cos\theta = \sqrt{(1 - x^2/a^2)} = \sqrt{(a^2 - x^2)}/a \text{ and}$$
$$\frac{y}{a} = \sin\theta\cos\phi + \cos\theta\sin\phi$$
$$= \frac{x}{a}.\cos\phi + \frac{\sqrt{(a^2 - x^2)}}{a}.\sin\phi$$

whence $y - x\cos\phi = \sqrt{(a^2 - x^2)}.\sin\phi$

and $x^2 - 2xy\cos\phi + y^2 = a^2\sin^2\phi.$

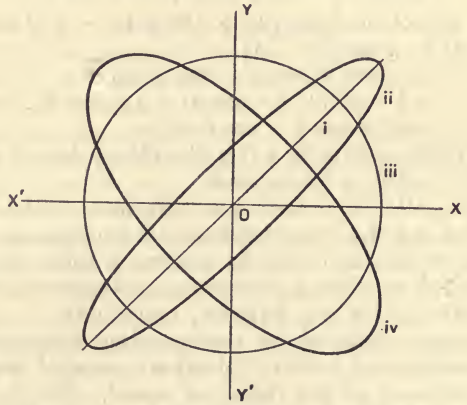


Fig. 129.

To turn the curve clockwise through 45° (**No. 27**) we substitute for x $(x - y)/\sqrt{2}$ and for y $(x + y)/\sqrt{2}$ and so obtain

$$x^2(1 - \cos\phi) + y^2(1 + \cos\phi) = a^2\sin^2\phi$$

or $x^2.\sin^2\frac{\phi}{2} + y^2.\cos^2\frac{\phi}{2} = \frac{a^2}{2}.\sin^2\phi = 2a^2\sin^2\frac{\phi}{2}.\cos^2\frac{\phi}{2}$

that is an ellipse whose semi-axes are

$$\sqrt{2}.\cos\frac{\phi}{2} \text{ and } \sqrt{2}.\sin\frac{\phi}{2}.$$

Fig. 129 shows how the form of the ellipse changes as ϕ increases. The curves marked (i), (ii), (iii), (iv) are those of **No. 29**.

In **No. 31** we have $x = a\sin\theta$, $y = a\sin 2\theta$, in **No. 32**

$y = a \cos 2\theta$, in No. 33 $y = a \cos 2\theta$. The elimination of θ is simple in each case. The first two of the three curves

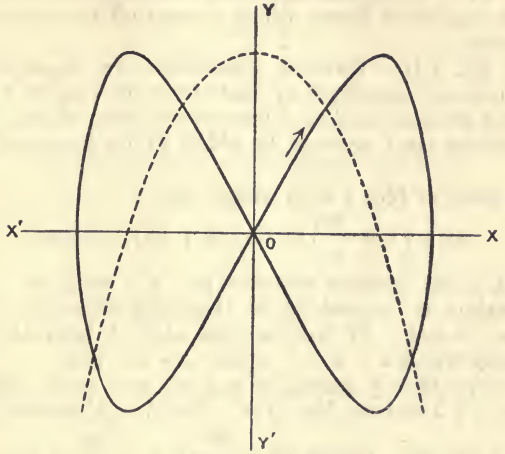


FIG. 130.

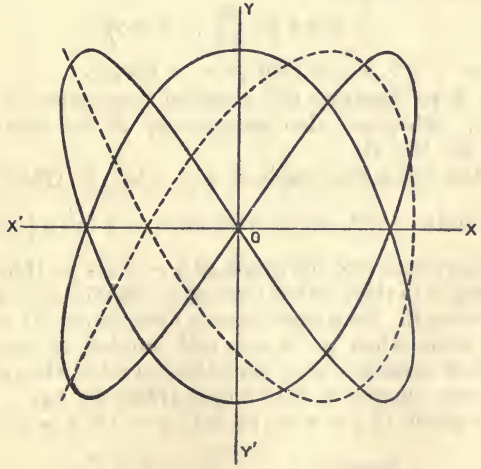


FIG. 131.

(**No. 34**) are shown in fig. 130. Fig. 131 gives the curves of **No. 35**, the broken being the one in which the phase difference is 45°. Text-books on sound usually give these and the more complicated forms which correspond to other ratios of frequency.

§ 7. Ex. CII. *Circular Functions.*—In these exercises the important transition is made from the idea of a sine or cosine of an *angle* to that of the sine or cosine of *any variable* x. Nothing need, perhaps, be added to the discussion in the text.

The proof of **No. 1** is, of course, that

$$\sin p\left(x + \frac{2\pi}{p}\right) = \sin(px + 2\pi) = \sin px.$$

In **No. 2** the formula becomes $y = a + b \sin(x - c)$. It has maxima at intervals of 2π, beginning where $x - c = \pi/2$ or $x = c + \pi/2$. It has minima also at intervals of 2π, beginning where $x - c = -\pi/2$ or $x = c - \pi/2$.

To solve **No. 5** assume $y = a \sin p(x + b)$. Then we have $a = 3{\cdot}2$ and (by No. 1) $p = 2\pi/2{\cdot}7$. Also when $x = 0$, $y = 3{\cdot}2 \sin pb$. Hence $pb = \frac{2\pi}{2{\cdot}7} . b = -\frac{2\pi}{3}$ or $b = -0{\cdot}9$. The formula is, therefore,

$$y = 3{\cdot}2 \sin \frac{2\pi}{2{\cdot}7}(x - 0{\cdot}9)$$

when $x = +2{\cdot}7$, $y = 3{\cdot}2 \sin \frac{4}{3}\pi = -1{\cdot}6\sqrt{3}$.

Nos. 8-10 illustrate the practical importance of the sine function. They are also anticipatory of the study of the tides in Ex. CX, E.

Fig. 132 gives the graph of $y = a \tan px$ (**No. 12**) with $a = 1{\cdot}5$ and $p = 0{\cdot}9$. Since $a \cot px = -a \tan p\left(x \pm \frac{\pi}{2p}\right)$ it can be converted into the graph of $y = a \cot px$ (**No. 14**) by (i) moving it to right or left through a distance $\pi/2p$ and then (ii) inverting it. Both functions are discontinuous; for $\tan px$ has no value when px is any odd number of right angles (positive or negative), while $\cot px$ has no value when px is zero or any even number of right angles (**Nos. 13, 14**). Fig. 133 gives the graph of $y = a \sec px$ with $a = 1{\cdot}6$, $p = 0{\cdot}8$. Since

$$a \operatorname{cosec} px = -a \sec p\left(x + \frac{\pi}{2p}\right)$$

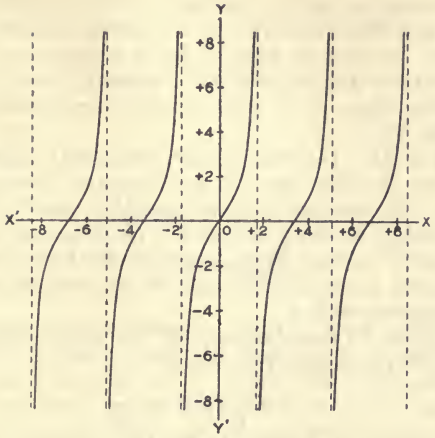


FIG. 132.

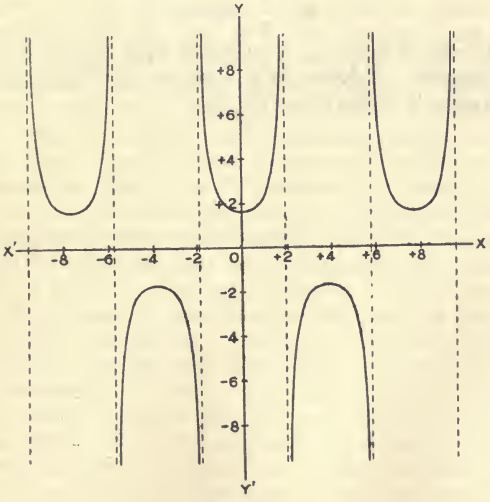


FIG. 133.

it can be converted into $y = a$ cosec px by (i) moving it $\pi/2p$ to the left and then (ii) inverting it. Both functions are discontinuous; for sec px has no value when px is any odd number of right angles (positive or negative), while cosec px has no value when px is zero or any even number of right angles (**No. 17**).

§ 8. Ex. CIII. *Inverse Circular Functions.*—This exercise travels over familiar ground and requires, therefore, little comment. The teacher is recommended to employ the continental notation arc sin x, arc tan x in preference to our insular $\sin^{-1} x$, $\tan^{-1} x$, etc. Since, however, the latter form is still generally used in examination papers the student should be made acquainted with it.

Fig. 132 can be transformed, by the usual inversion and rotation, into the graph (**No. 10**) of

$$y = \frac{1}{p} \cdot \text{arc tan } \frac{x}{a}.$$

The same treatment applied to fig. 133 converts it into the graph of

$$y = \frac{1}{p} \cdot \text{arc sec } \frac{x}{a},$$

a simple case of which is required in **No. 15** (ii).

The identities of **Nos. 29** and **30** are needed in connexion with Gregory's series (Ex. CX, C).

CHAPTER XLVIII.

WAVE-MOTION.

§ 1. *Preliminary Remarks.*—The general spirit in which the subject of wave-motion is to be attacked has been indicated in ch. XXXVIII., § 5. From the mathematical point of view wave-motion is a specially interesting concrete example of the dependence of one variable upon the simultaneous values of two others. Thus, in the familiar case of water-waves, the height of the water above the level of the undisturbed surface is a function both of the position of the point and of the time; that is to say it varies from point to point at the same moment and from moment to moment at the same point. In general, then, the study of waves plays, in connexion with the doctrine of the variation of a dependent variable with two independent variables, the part which the study of a moving point plays in connexion with the doctrine of dependence upon a single variable. In particular, since waves are, in most instances, periodic phenomena, they also afford excellent concrete instances of periodic functions. Apart from the circumstance that the phenomena of waves offer so powerful a stimulus and such effective guidance in an important department of mathematical thought they are, merely as phenomena, so familiar and impressive that they invite mathematical analysis and description as naturally as any other performances of nature or of man. Further, no one can be in a position to appreciate the most striking triumphs of physical science who has not given some attention to the elementary grammar of wave-motion. When to these is added the further consideration that the mathematical treatment required is, in its essential features, as simple as it is beautiful, it must be recognized that the inquiry is one into which the

ordinary student should no longer be debarred from entering merely because it does not form part of a somewhat amorphous tradition.

The theory of the subject is developed in Exs. CIV, CV, and CVI. The first is given to "progressive" and the second to "stationary" waves—phenomena whose special features are sufficiently described in the text. In these exercises we consider not only waves which can be described by a simple function of x and t, but also those which correspond to the formulæ based on a combination of single functions. In the third exercise an attempt is made, by very simple methods, to illustrate the famous process by which the great Fourier taught mathematicians how to analyse a given system of waves of any complexity. As an important addendum to these exercises the last section of the supplementary Ex. CX is devoted to the fascinating question of tidal prediction.

§ 2. Ex. CIV. *Progressive Waves.*—The chief phenomena of progressive waves are so familiar that it is not necessary to exhibit them in the class-room. Nevertheless the teacher who has a gift for experimenting will find that the mathematical analysis gains point and interest if prefaced by observations directed *ad hoc*. A length of narrow stair-carpet serves ideally well for the exhibition both of a single pulse and of a train of waves. Rough measurements of the velocity and (in the case of a train) of the wave-length and frequency are easily made and add to the clearness of the students' initial ideas. A length of india-rubber gas-tubing (which can be filled with sand) is more likely to be part of a laboratory equipment. A length of rope (within limits, the heavier the better) will do as a substitute.

As in ch. XLVII., § 2, the experiments intended to aid the analysis of the motion are for our purpose more important than those which exhibit it. The simple model described in the exercise (§ 2) should certainly be made by the teacher for exhibition, if not by each student. Its use brings out vividly the relation of the movements at certain points to one another and to the motion of the "wave". The following very easy device is also remarkably effective. Take a wooden rod, rather more than a yard long.[1] At intervals of an inch

[1] A broomstick serves the purpose perfectly well—or a length of 1 inch "dowell-rod".

along its length stick into it a series of hat-pins about 4 inches long, arranged in a uniform spiral. Thirty-seven pins should be used and should be arranged so that the spiral makes three complete turns. In this case the pins, if viewed along the length of the rod, are separated from one another by a uniform interval of 30°. The model should be held horizontally, at some distance from a light, in front of a screen of white card or paper and turned rapidly between the fingers. (The teacher with a constructive gift may devise a more elaborate arrangement for producing rotation, but, with a little practice, the use of the fingers is sufficient.) As the model is rotated the shadows of the heads upon the screen will exhibit beautiful progressive wave-motion in the manner described in the text in the last paragraph of § 2. It becomes quite evident that the progressive harmonic wave is in this case constituted by equal simple harmonic motions of the shadows of the heads, each of which differs in phase from the motions of adjoining shadows by a constant amount—here 30°.

The formulæ of **Nos. 1** and **2** are fundamental. They are simply applications of the general principle that if a graph is moved a distance d, parallel to the x-axis, $x - d$ must be substituted for x in the formula. In this case the shifting in time t is vt. Hence the formulæ $y = a \sin p(x \mp vt)$.

No. 3 is also extremely important and every student must master it. (The principle has, of course, already been studied in the preceding exercises.) (i) Let λ be the wave-length; then to increase x by λ must be to increase the "argument" $p(x - vt)$ by 2π. That is $p(x + \lambda - vt) = p(x - vt) + 2\pi$ or $p = 2\pi/\lambda$. (ii) Let T be the periodic time of the wave; then since one individual point goes through a complete cycle of movements while the wave-form advances through a length λ, we have $T = \lambda/v$, or $v/\lambda = 1/T$. (iii) But if the time for a complete cycle is T seconds the frequency, or number of complete cycles per second is $1/T = n$.

We pass to the examples on compound harmonic waves in division B. In **No. 15** the two component forms are

$$y = A \cos \phi \,.\, \sin \frac{2\pi}{\lambda} x$$

and $$y = A \sin \phi \,.\, \sin \left(\frac{2\pi}{\lambda} x + \frac{\pi}{2} \right) = A \sin \phi \,.\, \cos \frac{2\pi}{\lambda} x.$$

The resultant wave-*form* is, therefore,

$$y = \mathrm{A}\cos\phi \,.\, \sin\frac{2\pi}{\lambda}x + \mathrm{A}\sin\phi \,.\, \cos\frac{2\pi}{\lambda}x$$

$$= \mathrm{A}\sin\left(\frac{2\pi}{\lambda}x + \phi\right).$$

The resultant wave-*motion* is

$$y = \mathrm{A}\sin\left\{\frac{2\pi}{\lambda}(x - vt) + \phi\right\}.$$

No. 24 sets the problem of combining waves of the same length in its most general form. We have for the resultant

$$y = a\sin px + b\sin(px + \phi)$$

$$= (a + b\cos\phi)\sin px + b\sin\phi \,.\, \cos px.$$

Assume (as in No. 18) that

$$(a + b\cos\phi) = \mathrm{A}\cos\phi' \text{ and } b\sin\phi = \mathrm{A}\sin\phi'$$

then

$$\mathrm{A} = \sqrt{\{a^2 + 2ab\cos\phi + b^2\}} \text{ and } \tan\phi' = \frac{b\sin\phi}{a + b\cos\phi},$$

and

$$y = \mathrm{A}\cos\phi' \,.\, \sin px + \mathrm{A}\sin\phi' \,.\, \cos px$$

$$= \mathrm{A}\sin(px + \phi').$$

In **No. 26** since $\frac{2\pi}{\lambda} = \frac{\pi}{10}$, $\lambda = 20$. The unbroken thick line in *Exercises*, II, fig. 97, represents one period of the resultant—the vertical scale being considerably exaggerated.

The formulæ of **No. 27** substitute for $2\sin\frac{\pi}{5}x$ (i) $- 2\cos\frac{\pi}{5}x$, (ii) $+ 2\cos\frac{\pi}{5}x$. *Exercises*, II, fig. 98, shows one wave-length in the latter case. Fig. 134 of this book shows one wave-length of the resultant in **No. 28**. The component wave-lengths are 8 and 12 so that the resultant wave-length is their least common multiple, 24.

§ 3. Ex. CV. *Stationary Waves.*—To *produce* stationary waves the teacher may have recourse to the length of stair-carpet or a substitute. A strip of carpet, 10 or 12 feet long, can very easily be made to vibrate with two internodal segments. The effect is almost equally well produced with a length of rope or tubing weighted with sand. A long narrow trough, fitted with a glass side, is found in many laboratories and can be used for experiments on stationary water-waves. It should be used as described in **No. 17**. The velocity of water-waves whose length is large compared with the depth is given by

the formula $v = \sqrt{gh}$, where h is the depth in feet or centimetres and $g = 32$ in the first case and 981 in the second. It is not difficult to verify this formula by counting the frequency of the uni-nodal vibrations and using the relation $n = v/(2l)$.

For the *analysis* of stationary wave-motion pieces of thick iron wire about 30 inches long, bent and used as described in **Nos. 3, 4, 5**, give striking results—except, perhaps, for the case $y = a \sin \pi x/l$. Consideration of the rotated wire leads easily to the fundamental formula of **No. 3**. It is evident that the point P_1 moves in a circle of radius $a \sin \pi x/l$. If the time is counted from a moment when the plane of the wire is vertical the projection of this radius upon the wall is, at time t,

$$y = a \sin \frac{\pi}{l} x \,.\, \cos 2\pi nt.$$

Similar arguments apply in **Nos. 4, 5**.

When the proper moment has come each student should certainly perform or see the experiment described in **No. 22**.

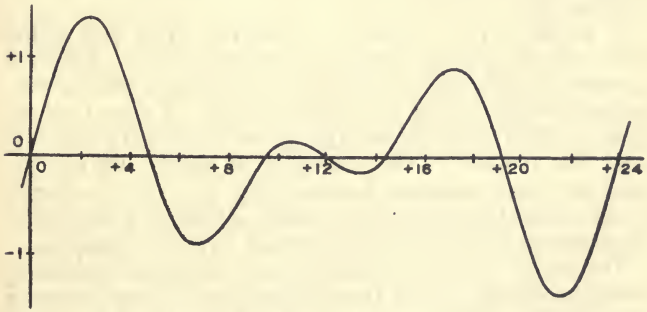


FIG. 134.

Nos. 11-16 are extremely important but should give no difficulty. The difference between the two formulæ of **No. 16** is that in the former the two waves are supposed to be $y = \mathrm{A} \sin \frac{2r\pi}{\lambda} (x \pm vt)$ and in the latter $y = \mathrm{A} \cos \frac{2r\pi}{\lambda} (x \pm vt)$. In the first case y is zero when x and t are zero; in the second y has then its maximum value A. In other words the difference is simply one of origin. The necessity for

the alternative formula comes into view in **No. 17**; for in the experiment with the trough the ends are obviously the places of most disturbance. As is seen by putting $x = 2{\cdot}5$, the node is half way along the trough. A very interesting example of a natural water-motion similar to that of **No. 17** is afforded by the curious phenomena observed on Lake Geneva and called *seiches*. The lake appears at times to vibrate about a nodal line drawn across its length. Taking the average depth to be 115 metres $v = \sqrt{981 \times 115 \times 100}$ $= 33{\cdot}5$ m./sec. $\div 2$ kilometres per minute. Also $l = 70$ kilometres. In accordance with these data the vibrations of the surface should have a period of $2l/v = 70$ minutes, and should be describable by the formula

$$y = a \cos \frac{\pi x}{70} \cdot \cos \frac{\pi t}{35}.$$

The formula is found to agree curiously well with the facts, a varying from a few millimetres to as much as a metre or more. Interesting details (with references to the original sources) are given in G. H. Darwin's *Tides and Kindred Phenomena*.

§ 4. Ex. CVI. *Harmonic Analysis.*—The first article deals with the analogy between a series of integral powers of x, regarded as the natural means of expressing a class of non-periodic functions, and a "Fourier series" of sines (and cosines) as the symbolic idiom appropriate for the expression of a periodic function. The analogy is, of course, not perfect; there are many important functions (e.g. $y = \sqrt{x}$, $y = \log x$), which cannot be expressed as a series of integral powers of x. Nevertheless, it is very instructive and is worth emphasis.

The rest of the exercise deals with the practical problem of resolving a given periodic function (or curve) into its harmonic components. The reader who is not already familiar with Fourier's theorem should make himself quite clear as to what it asserts. Let him turn, for example, to *Exercises*, II, figs. 97 and 98. Each of these diagrams exhibits a pattern which is supposed to be repeated innumerable times to the right and to the left, each repetition occupying a length λ. In fig. 97 the dotted line and the line drawn in dots and dashes show how the firm-lined curve was built up. The two components were a "fundamental" sine-curve of a certain amplitude whose axis-length is identical with that of the

wave—that is λ, and another sine-curve whose length is $\lambda/2$ and whose amplitude is one-half of that of the former curve. It is evident that if there had been other components, sine-curves of lengths $\lambda/3$, $\lambda/4$, etc., the wave-form would have shown modifications which could be made indefinitely varied by varying the selection of the subsidiary sine-curves and their amplitudes. Much the same can be said of fig. 98 with the important difference that the curve in that case is the resultant of a simple sine-curve of length λ and a simple *cosine*-curve of length $\lambda/2$. Thus this figure suggests that periodic curves of increasing complexity can be built up by taking as the fundamental component either a sine-curve or a cosine-curve of length λ and adding to it either sine-curves or cosine-curves of lengths $\lambda/2$, $\lambda/3$, $\lambda/4$, etc.

Now Fourier's theorem asserts two things. The first is that the preceding statement can be made conversely; that is, *any* repeated pattern of length λ can be produced by adding to a certain fundamental sine- or cosine-curve of that length sine- or cosine-curves of the proper amplitudes whose lengths are $\lambda/2$, $\lambda/3$, $\lambda/4$, etc. Thus if y is the ordinate of any periodic curve of length λ or the value of any periodic function whose period is λ we have

$$y = a_0 + a_1 \sin px + a_2 \sin 2px + a_3 \sin 3px + \ldots$$
$$+ b_1 \cos px + b_2 \cos 2px + b_3 \cos 3px + \ldots$$

where $p = 2\pi/\lambda$. (The constant a_0 must be inserted to meet the case in which the x-axis is not identical with the common axis of the harmonic curves.)

The second thing asserted by Fourier's theorem is that by a certain process any of the unknown amplitudes can be determined at will. It is usual to express and to demonstrate the method by means of the integral calculus. As examples **Nos. 1-14** show, such heavy weapons are unnecessary. The method can be made to depend simply upon the obvious fact that the total area of either a complete sine-curve or a complete cosine-curve is zero—because half of it lies on each side of the axis. In **Nos. 6-12** a very important consequence is seen to follow from this property. Imagine a solid whose "plan" is bounded by the curve $\sin rpx$ (or $\cos rpx$) and the axis, and whose "elevation" is bounded by the same axis and the curve $a \sin spx$ (or $a \cos spx$) r and s being any two positive integers. Then—taking into account the ordinary conventions about the sign of an area—it can at once be

shown that this volume is always zero if r and s are different and always $al/2$ when they are the same—l being the axis-length of the curve under analysis. Division B of the exercise shows how this striking property can be used to "pick out" the amplitude of any specified component of the given curve.

The student is directed in No. 6 and elsewhere to draw his plans and elevations and to set them at right angles to one another. The teacher will find it useful to carry these instructions further and actually to build up one or two of the described solids in plasticine or other plastic material. The solid based upon the curves $a \sin px$ and $\sin 3px$ is a suitable one for the purpose. The curves are best cut out of thin sheet zinc (though cardboard will suffice), any convenient value being assigned to a. The two positive volumes should be built up of blue plasticine and the negative volume of red. The segments should then be removed and placed, the positive segment upon one pan of a balance, the two negative segments upon the other. If they have been carefully modelled it will be found that the negative segment just balances the two positive segments—that is, that the total volume is zero.

In calculating the amplitudes of the components it is well to begin by making a table of the measured heights of a series of equidistant ordinates of the given curve. In the examples given it will be sufficient to take these so that they represent phase-differences of 15° in the first component, 30° in the second component, etc. This will imply the measurement of thirteen ordinates in the length AB′ of **No. 15** and twenty-five ordinates in AB of **No. 18**, both terminal ordinates being included. The following table shows the data and the method of working in **No. 15**. The first column gives the numbers of the ordinates. The second column, headed "y," gives the heights of the eleven intermediate, and one-half the height of the initial and final, ordinates. In the present instance, since both the terminal ordinates are zero, the halving does not make itself apparent. In the third column we have the product of each of the ordinates by $\sin 15n°$, n being the number of the ordinate. These products are the areas of the successive sections (but the half areas of the end sections) of the solid of which one half of the given curve is the elevation and the curve $y = \sin \pi x/l$ is the plan. The two columns headed "$y \,.\, \sin 30n°$" contain the products which measure the

successive cross-sectional areas (but half the end-areas) of the solid in which the elevation is as before but the plan is the curve $y = \sin 2\pi x/l$. It should be noted that the positive and the negative products are set down separately. All these products were obtained in a very few minutes by means of a slide-rule. When each member of the class measures a single ordinate and computes the corresponding products the work is done very quickly.

n	y mms.	y . sin $15n°$.	y . sin $30n°$. +	−
0	0·0	0·00	0·00	
1	13·0	3·46	6·50	
2	23·0	11·50	19·93	
3	29·5	20·90	29·50	
4	31·5	27·40	27·40	
5	29·5	28·60	14·75	
6	24·0	24·00		0·000
7	17·5	16·90		8·750
8	10·3	9·18		8·925
9	5·0	3·54		5·000
10	2·0	1·00		3·714
11	1·0	0·26		0·500
12	0·0	0·00		0·000
		+ 146·74	+ 98·08	− 24·89
			− 24·89	
			+ 73·19	

Calling either of these totals S we have by Simpson's Rule that the volume V is $Sl/12$, for $l/12$ is the common interval between the sections. But the average area, $\bar{A}$, is V/l; hence $\bar{A} = S/12$. Thus the rule, $a = 2\bar{A}$, becomes in the present instance $a = S/6$. We deduce, therefore, that $a_1 = 2{\cdot}45$ cms., $a_2 = 1{\cdot}22$, and that $y = 2{\cdot}45 \sin \pi x/l + 1{\cdot}22 \sin 2\pi x/l$. These results are in practically perfect agreement with the facts.

In **No. 18**, if the instructions given in the text are followed, each student will make a table consisting of two columns answering to the first two in the foregoing table and a pair of columns answering to the last two. In the case of the students who are seeking b_2, the amplitude of the assumed component $b_2 \cos 4\pi x/l$ (l being now = AB = λ), the table will be as follows :—

n	y	$y \cdot \cos 30n°$.	
	mms.	+	−
0	+ 6·25*	6·25	
1	17·00	14·85	
2	18·5	9·25	
3	17·2	0·00	
4	14·8		7·80
5	12·2		10·52
6	11·9		11·90
7	12·2		10·52
8	14·9		7·45
9	17·3		0·00
10	18·4	9·20	
11	17·8	15·42	
12	13·5	13·50	
13	+ 5·0	4·33	
14	− 5·0		2·50
15	16·0		0·00
16	26·0	13·00	
17	32·0	27·75	
18	35·5	35·50	
19	34·5	29·80	
20	27·6	13·80	
21	20·0	0·00	
22	7·3		6·32
23	0·0		0·00
24	+ 5·65*	5·65	
		+ 98·3	
		− 57·0	− 57·0
		+ 141·3	

* Half the ordinate.

In this case $V = Sl/24$, $\bar{A} = V/l = S/24$, and $b_2 = 2\bar{A}$. Hence $b_2 = S/12 = 11{\cdot}8$ mms. The correct result is 12 mms. The slight difference is due partly to uncertainties in reading the height, partly to the draughtsman's errors in reproducing the graph from the original drawing.

No. 20 differs from the former examples in that the ordinates can (and should) be determined by calculation. In order to include the ordinate at P it will be best to suppose the ordinates drawn at intervals of 4 inches—that is, so as to give phase-differences of 12° in the case of the first component. From A to P and from P to B the heights are given respectively by the formulæ

$$y = \frac{x}{24} \times 2 \text{ and } y = \frac{60 - x}{36} \times 2$$
$$= \frac{x}{12} \qquad = \frac{10}{3} - \frac{x}{18}.$$

Including the cases $x = 0$ and $x = 60$ we have the following sixteen ordinates of which the first and last may be supposed to have been halved :—

$$0, \tfrac{1}{3}, \tfrac{2}{3}, 1, 1\tfrac{1}{3}, 1\tfrac{2}{3}, 2, \tfrac{16}{9}, \tfrac{14}{9}, 1\tfrac{1}{3}, \tfrac{10}{9}, \tfrac{8}{9}, \tfrac{2}{3}, \tfrac{4}{9}, \tfrac{2}{9}, 0.$$

When these numbers are multiplied in order by sin 0°, sin 12°, sin 24° . . . sin 180°, we obtain the fourteen whole areas and the two semi-areas needed to find by Simpson's

Rule the volume of the solid by which the value of a_1 is to be determined. Proceeding as before we have $S_1 = + 12{\cdot}09$, $V_1 = 12{\cdot}09l/15$, $\bar{A}_1 = 12{\cdot}09/15$ and $a_1 = 2\bar{A}_1 = 1{\cdot}61$.

When the same sixteen ordinates are multiplied in order by sin 0°, sin 24°, sin 48°, . . . sin 360°, the total, S_2, is 1·89 and A_2 is 0·126. Hence

$$a_2 = 2\bar{A}_2 = 0{\cdot}25.$$

For the third component the numbers must be multiplied by sin 0°, sin 36°, sin 72° . . . sin 540°, and for the fourth component by sin 0°, sin 48°, sin 96°, . . . sin 720°. The corresponding values of S_3 and S_4 are $-0{\cdot}855$ and $-0{\cdot}799$. We have, therefore,

$$a_3 = 2\bar{A}_3 = -0{\cdot}113 \qquad \text{and} \qquad a_4 = 2\bar{A}_4 = -0{\cdot}106.$$

As far as our analysis has gone we have found, then, that the formula of the curve is

$$y = 1{\cdot}61 \sin \pi x/60 + 0{\cdot}25 \sin \pi x/30 - 0{\cdot}11 \sin \pi x/20 - 0{\cdot}11 \sin \pi x/15.$$

The formula asked for in **No. 20** is

$$y = 1{\cdot}61 \sin \pi x/60 \,.\, \cos 2\pi t/T + 0{\cdot}25 \sin \pi x/30 \,.\, \cos 4\pi t/T - 0{\cdot}11 \sin \pi x/20 \,.\, \cos 6\pi t/T - 0{\cdot}11 \sin \pi x/15 \,.\, \cos 8\pi t/T$$

If we substitute $t = 0$ we obtain the initial form of the string; if we put $t = l/2v$, the form when the fundamental component

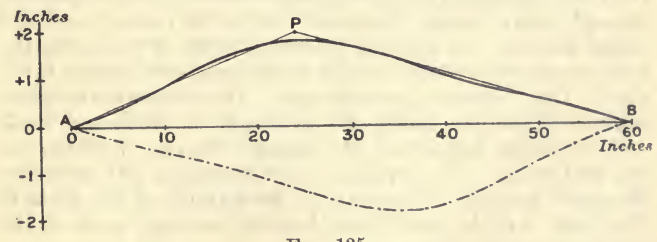


Fig. 135.

has completed half a vibration. The shapes which these formulæ assign to the cord **No. 21** are shown in fig. 135, the vertical displacements being exaggerated five times.

Fig. 136 is the graph of **No. 22** (ii).

§ 5. *The Tides.* Ex. CX, E.—It is unfortunate that any-

thing like a full statement of the results of tidal analysis would be utterly beyond the scope of an elementary book. Nevertheless the subject is so interesting and important that it could not be omitted from a section dealing with periodic functions. Division E of the supplementary Ex. CX deals with the one topic which is amenable to simple treatment —namely, the dependence of the time and depth of high water upon the angular distance between the sun and moon when the latter body is on the meridian. The data of the table on p. 317 are taken from the papers by Lubbock in *Philosophical Transactions*, 1831-7. It must not be forgotten that they are the means of hundreds of observations taken when the sun and moon were occupying all the variety of positions possible to them at different times in the cycle of their movements. Thus predictions based upon the table may disagree, sometimes to the extent of a quarter of an hour or so, with

FIG. 136.

the actual times of high water. Lubbock in the same papers examined in great detail the way in which the means must be corrected according to the declination of the moon and of the sun, the moon's distance from the earth, etc.

The theoretical formulæ of **Nos. 49** and **50** were obtained from the hypothesis that the tides are caused by the formation of a spheroid of water whose axis follows the moon round and round the sky. If the axis pointed directly to the moon the tides would everywhere be high when the moon is on the meridian. That, however, is not the case. The axis makes a certain angle with the line joining the centres of the earth and moon and this angle varies with the angular distance between the sun and moon. The number in the table of p. 317 measures the angle in time. It appears in the formula of No. 49 as θ. The angle λ in the same formula is the average value of the angle for a given port.

The angle ϕ which also appears in the formula is the angular distance between the sun and moon, and is measured by the interval between their transits. It increases at the rate of about 48 m. or 12° a day. The fact that the constant α is, for London, 2 h. or 30° is interpreted as meaning

that the tide which follows a certain transit of the moon is, in reality, determined, not by the present position of the moon with regard to the sun but by its position $30°/12° = 2\frac{1}{2}$ days ago. This is, then, the time which the tidal wave raised in the vast area of the Pacific Ocean takes to reach London.

The formula of **No. 50** is derived from the same hypothesis by theoretical considerations. It gives the height of the pole of the spheroid above its mean surface.

The subject of the tides has a considerable literature much of which can be consulted, for teaching purposes, with advantage. The original papers of Lubbock, together with the extremely lucid contributions by Dr. Whewell, which appeared in *Philosophical Transactions* about the same time, will be found very illuminating by the teacher who has access to them. They are much more within the scope of the mathematician of ordinary powers and the ordinary amount of leisure than the monumental papers of Sir G. H. Darwin. On the other hand Darwin has published in *The Tides and Kindred Phenomena* (John Murray) a book which, though "popular," is full of original and interesting information upon this subject. The article on the tides in the *Encyclopædia Britannica* was also written by Darwin and gives an easily accessible summary of the development of the subject. Lastly, Lord Kelvin included important lectures upon waves, ripples, and tides in the well-known volumes which should have a place in every school library.

CHAPTER XLIX.

DIFFERENTIAL FORMULÆ OF THE CIRCULAR FUNCTIONS.

§ 1. Ex. CVII. *Differential Formulæ for Sine and Co-sine.* —This topic might have been considered at an earlier point —for example, in connexion with simple harmonic motion. It is withheld, partly in order not to interrupt the main stream of the argument and partly because its introduction at the point chosen helps to develop further the analogy between non-periodic and periodic functions. Differential formulæ for sin *rpx* and cos *rpx* are now needed to answer questions about periodic functions exactly in the same way as differential formulæ for x^r were needed to answer questions about the non-periodic functions of Part I. This is the point of view from which the subject-matter of Ex. CVII is presented in the introductory article.

In the actual deduction of the differential formulæ (**Nos. 6-9**) the student is led to examine rather carefully the exact significance of the various steps. This increasing "rigour" is appropriate to a higher stage of mathematical experience; and it prepares the student for the later treatment based on the concept of limit (Ex. CXII).

The examples in divisions A and B contain nothing that calls for comment. In division C the newly found formulæ are applied to questions of simple harmonic motion. The questions are quite simple but the answers have most important applications in physics.

The solution of **No. 25** is:—

$$\frac{\delta x}{\delta t} = 2\pi n a \cos 2\pi n t$$
$$= 2\pi n \sqrt{a^2 (1 - \sin^2 2\pi n t)}$$
$$= 2\pi n \sqrt{(a^2 - x^2)}.$$

No. 29 is of fundamental importance but still easier:—

$$\frac{\delta x}{\delta t} = 2\pi na \cos(2\pi nt - \phi)$$

$$\frac{\delta^2 x}{\delta t^2} = -(2\pi n)^2 a \sin(2\pi nt - \phi)$$

$$= -4\pi^2 n^2 . x$$

§ 2. Ex. CX, A. *Expansions of Sine and Cosine.*—The first division of the supplementary exercise begins with examples in which differential formulæ are applied to various practical problems. In connexion with these the student has to face a difficulty which sometimes gives trouble. The differential of $\sin\theta$ is $\cos\theta.\delta\theta$ only if θ is measured in radians. If (as will commonly be the case in practical problems) he is dealing with a $\sin\alpha$ in which α is measured in degrees, he cannot assume that the differential is $\cos\alpha.\delta\alpha$. It is first necessary to convert the α degrees into $\pi\alpha/180$ radians. We then have

$$\delta\left(\sin\frac{\pi\alpha}{180}\right) = \frac{\pi}{180} . \cos\frac{\pi\alpha}{180} . \delta\alpha = \frac{\pi}{180} . \cos\alpha.\delta\alpha.$$

Nos. 6, 7, 12 should be treated carefully. They are important as laying a concrete foundation for the study of the derivative of a product of functions and of partial differentiation in Section VIII. The "reason" expected in No. 6 is as follows: The errors in a due to errors in β, a and b, are, by hypothesis, small. If another number used in calculating one of these errors is itself subject to a small error it will produce a small error in the estimation of the small error in question. But by the very definition of "small" a small error in a small error is unrecognizable. Hence the errors in a due to faulty measurements in β, a and b can be calculated independently of one another and the total error will be their sum.

In **Nos. 17-20** the subject of the expansion of the sine and cosine is resumed from Ex. XCIX, and completed upon a more satisfactory logical basis.[1]

§ 3. Ex. CX, B. *Differential Formula for tan x.*—This differential formula is important chiefly because it is needed in connexion with Gregory's series for π. For this reason it

[1] The author owes the beautiful central idea of the method to a friend who tells him that it is current in Cambridge but has not enabled him to attribute it to its inventor.

is considered in the supplementary exercise. **Nos. 21, 22** indicate a sound method of deduction. By making h small enough tan h may be made equal to h to any desired degree of approximation and so small that the factor $(1 - \tan x \,.\, \tan h)$ becomes as nearly equal to unity as we please. Since with further diminution of h the value of tan h is simply proportional to h, we may write

$$\delta y = \frac{\delta x}{\cos^2 x}$$
$$= \sec^2 x \,.\, \delta x.$$

This result, in its simple form, holds good for tan θ only when θ is measured in radians. If the angle is measured in degrees we have $\delta(\tan \alpha) = \frac{\pi}{180} \sec^2 \alpha . \delta\alpha$. This consideration has to be applied in all the practical examples.

§ 4. Ex. CX, C. *The Calculation of* π.—This section contains examples of the calculation of π which have great historical importance. The methods all had their origin in the discussions which sprang from the publication of Wallis's *Arithmetica Infinitorum* in 1655. Gregory communicated his series for arc tan x in a letter to his correspondent Collins (1671) but gave no proof. There is little question that he reached it by the process indicated in **Nos. 31-3**.

The teacher who wishes to come to close quarters with the minds of the men who laid the foundations of modern mathematics will find an invaluable collection of documents in Maseres' *Scriptores Logarithmici* to which reference has already been made (p. 468). The statement about Machin's series in **No. 37** condenses into a sentence an interesting paper by Hutton, the author of the famous introduction to logarithmic tables.

CHAPTER L.

THE HYPERBOLIC FUNCTIONS.

§ 1. *Value of the Subject.*—The hyperbolic functions cannot be said to be a subject of fundamental importance; nevertheless there are good reasons for giving them a not inconspicuous place in a general course. The chief of these is the beautiful parallelism which exists between the circle and the circular functions on the one hand, and the rectangular hyperbola and the hyperbolic functions on the other. In the next place, the discovery of simple algebraic expressions for the values of $\sinh x$ and $\cosh x$ fortifies the student's notions of $\sin x$ and $\cos x$ as functional relations which do not necessarily imply any connexion with angular measurement. Again, as was remarked before, the hyperbolic functions wait, so to speak, upon the circular functions in many regions of physics. Thus, the change of sign produced by increasing one term in a differential formula relatively to another may imply that a body's motion is transformed from periodic oscillations described by a circular function into a non-periodic movement described by a hyperbolic function. Finally, we have seen that there is a useful analogy between a power-series of x and a sine or cosine series—the one playing with regard to many important non-periodic functions the part which the other plays in connexion with periodic functions. The discovery of the hyperbolic functions shows that the analogy can be carried into a field—that of the exponential functions—where the power-series ceases to be the natural instrument of analysis and statement. Indeed it shows that the analogy between periodic functions and the non-periodic functions of this field is the closest of all.

§ 2. Ex. CVIII. *The Hyperbolic Sine and Cosine.*—The argument of the exercise may be summarized as follows: Pt. I, Ex. LXV, Nos. 7-12 brought out the fact that all ellipses

can be supposed derived from the circle and all hyperbolas from the rectangular hyperbola by an identical process. Thus there is a striking correspondence between the curves $x^2 + y^2 = a^2$ and $x^2 - y^2 = a^2$, each taken as a whole. It is reasonable, therefore, to inquire whether there is not an equally striking correspondence between the individual points of the two curves. In **Nos. 7-12** two distinct principles of correspondence are suggested and are found to lead to an identical conclusion: namely, that the point $(a \sec \theta, a \tan \theta)$ on the hyperbola may be held to correspond to the point $(a \cos \theta, a \sin \theta)$ on the circle. In this way it is possible, then, to "pair off" any point on the one curve with a point on the other—with the exception (**No. 10**) that the points where the circle cuts the y-axis have no partners on the hyperbola.

The next question concerns the numbering of the corresponding points. Consider *Exercises*, II, fig. 99. For P, a point on the circle, θ, which measures the angle POA, is the natural numerical label. The point P′ on the hyperbola *might* also be labelled θ in virtue of its correspondence with P, but it would obviously be better to seek some principle of denumeration which may be applied without direct reference to the point on the circle. Two such principles readily offer themselves. The numerical value of θ is a measure (i) of the arc AP and also (ii) of the sectorial area AOP. It would be natural, therefore, to label P′ either (i) with a number which measures the hyperbolic arc AP′, or (ii) with one which measures the hyperbolic area AOP′. There is no *a priori* reason for choosing one of these rather than the other but there proves to be a strong practical reason for selecting the latter. It is not possible [1] to express the length of the arc AP′ definitely in terms of familiar functions of θ, but it is comparatively easy so to express the area AOP′. It is proved in **Nos. 13-16** that just as the area AOP can be written as $\frac{1}{2}a^2\theta$ so the area AOP′ can be written $\frac{1}{2}a^2u$ where the number u is calculated from θ by the relation

$$u = \log \tan \left(\frac{\theta}{2} + \frac{\pi}{4} \right).$$

By this formula it is possible to obtain a numerical label for any point P′ by considering merely its position on the hyperbola. For since its ordinate is $y = a \tan \theta$, we can determine

[1] The attempt to do so carries us into the difficult region of "elliptic functions".

θ by the relation $\theta = \text{arc tan}\,(y/a)$ and then proceed to calculate u without referring directly to the circle $x^2 + y^2 = a^2$.

Now if we consider a point P on the circle which is determined by a certain value of θ, the abscissa and ordinate of P are $a\cos\theta$ and $a\sin\theta$. In virtue of the general analogy between the curves it is appropriate to give the names *hyperbolic cosine* and *hyperbolic sine* of u to the functions which give the coordinates of the point P′ determined by a certain value of u. In virtue of these definitions we write

$$x = a\sec\theta = a\cosh u \text{ and } y = a\tan\theta = a\sinh u.$$

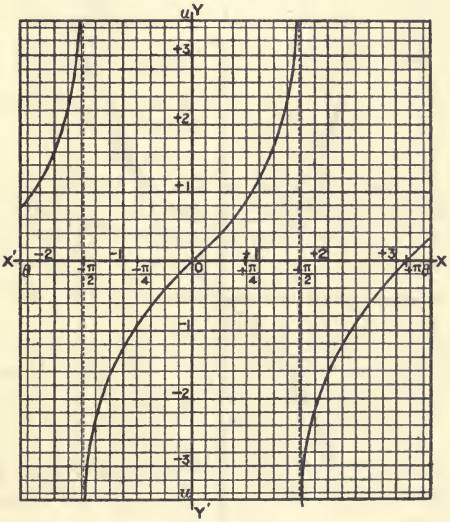


FIG. 137.

In the remaining examples the properties which flow from these definitions are studied. The graph of **No. 21** should be drawn with some care since it may be used in later examples. It is represented in fig. 137. The graph brings out (i) that u is a single-valued periodic function of θ; (ii) that it is discontinuous, since there are no values of u when $\theta = (2n+1)\frac{\pi}{2}$; (iii) θ is a many-valued function of u.

Fig. 138 shows the graphs of sinh n (S), cosh n (C) and tanh n (T) (**Nos. 22, 23**).

§ 3. Ex. CIX. *The Hyperbolic Functions.*—Just as the symbolism $y = \sin x$ may mean simply a connexion of a certain kind between an endless series of pairs of numbers, the original dependence of the function upon the properties of triangles being dropped out of sight, so $y = \sinh x$ and $y = \cosh x$ may be regarded as symbolizing connexions between numbers without reference to the rectangular hyperbola. The freeing of the hyperbolic functions from their

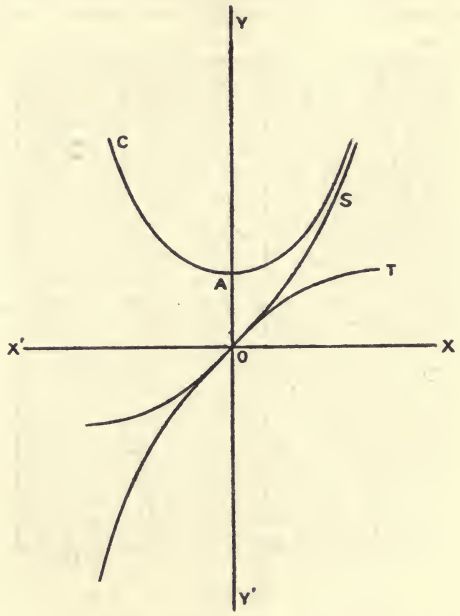


FIG. 138.

dependence upon the properties of a particular curve is best effected by showing that the values of $\sinh x$ and $\cosh x$ can be calculated directly by means of the exponential expressions $(e^x - e^{-x})/2$ and $(e^x + e^{-x})/2$. The demonstration of these equivalences is the subject of division A of the exercise. The consequent generalization of the idea of hyperbolic functions is studied in division B.

The examples from No. 11 onwards extend the analogy be-

tween the circular and the hyperbolic functions to their differentials. Thus in **No. 11** we have :—

$$y = \sinh x = \frac{e^x - e^{-x}}{2}$$

$$\therefore \frac{\delta y}{\delta x} = \frac{e^x + e^{-x}}{2} = \cosh x. \quad \text{[Ex. LXXXIII, No. 6.]}$$

Similarly in **No. 13** :—

$$y = a \sinh px = a \,.\, \frac{e^{px} - e^{-px}}{2}$$

$$\therefore \frac{\delta y}{\delta x} = ap \,.\, \frac{e^{px} + e^{-px}}{2} = ap \,.\, \cosh px$$

$$\text{and } \frac{\delta^2 y}{\delta x^2} = ap^2 \,.\, \frac{e^{px} - e^{-px}}{2} = ap^2 \,.\, \sinh px.$$

In the same way if $y = a \cosh px$

$$\frac{\delta^2 y}{\delta x^2} = ap^2 \cosh px.$$

From these results it follows (**No. 18**) that, if a point has any one of the motions $x = a \sinh pt$, $x = b \cosh pt$ or $x = a \sinh pt + b \cosh pt$, its "acceleration" is subject to the law

$$\frac{\delta^2 y}{\delta t^2} = p^2 x.$$

Thus we reach the striking and important conclusion that if the acceleration of a moving point is proportional to its distance from a fixed point the distance will itself be either a circular or a hyperbolic function of the time. It will be the former if the acceleration is $-p^2x$ and, therefore, is always directed back towards the origin. It will be the latter if the acceleration is $+p^2x$ and is, therefore, always directed away from the origin. In the former case the motion will be periodic or vibratory, in the latter case non-periodic or non-vibratory.

Nos. 22-4 bring out another striking and useful instance of the general analogy between the two kinds of functions. In No. 22 we have

$$y = a(\cosh px^2 - \sinh px^2)$$
$$= a \,.\, e^{-px^2}$$

Since the index of e is here necessarily negative the greatest value of y is a (corresponding to $x = 0$). The other values of y fall off symmetrically on both sides of the y-axis but

never reach zero value. The graph is, therefore, the single symmetrical "hump" depicted in the upper part of fig. 139. In this figure $a = 1$ and $p = 0{\cdot}1$ as in No. 24 (i). Now let the hump move to the right with velocity v. Then we have a single wave, involving all points of the axis, and describable by the formula of **No. 23**, namely

$$y = a\{\cosh p(x - vt)^2 - \sinh p(x - vt)^2\}.$$

A combination of such forms will produce a complex single wave which can be compared with the complex repeated waves obtained by combining harmonic waves. The lower part of fig. 139 shows, for example, the compound single wave of **No. 24** (i). Thus, just as the circular functions are the

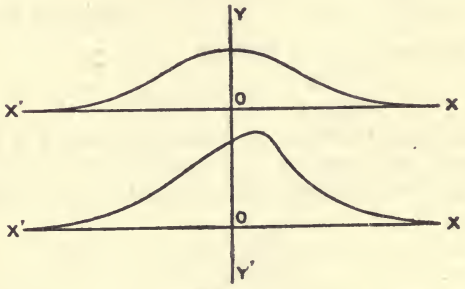


FIG. 139.

idiom most appropriate to the description of a train of identical waves, so the hyperbolic functions offer the most natural means of describing the single wave.

In Section IX the function e^{-px^2} will be found to play a most important part at every stage of the discussion. Nos. 22-4 may be regarded as preliminary to its use there.

§ 4. Ex. CX, D. *The Gudermannian Functions.*—When the student has learnt that $\delta(\tan x) = \sec^2 x \,.\, \delta x$ he is in a position to return to the study of the function of which the correspondence between points on a circle and its related rectangular hyperbola is the most important concrete instance. This function, symbolized in the general form

$$y = \log \tan \left(\frac{x}{2} + \frac{\pi}{4}\right),$$

can be regarded as the inverse of another function, or connexion between numbers, of the form

$$y = 2 \arctan e^x - \frac{\pi}{2}.$$

The latter is called the "Gudermannian function of x" and may, therefore, be written concisely as

$$y = \mathrm{gd}\, x.$$

The best corresponding symbolism[1] for the inverse Gudermannian function will be

$$y = \arg \mathrm{gd}\, x$$

though English writers generally write it

$$y = \mathrm{gd}^{-1} x.$$

By considering the correspondence between P and P′ in *Exercises*, II, fig. 99, it is easy to show that if $y = \arg \mathrm{gd}\, x$ then $\delta y = \sec x \,.\, \delta x$ (**Nos. 42-3**). The practical importance of this discovery constitutes the chief claim of the Gudermannian functions to a place in the course. The student is reminded of the principle by which he sought in Ex. LXXXVII to fix the positions of the parallels of latitude in the Mercator net. To apply that principle he had formerly to be contented with a graphic method. He is now able to see that y, the equatorial distance of a given parallel in the Mercator net, is connected with x, the equatorial distance of the parallel on the corresponding geographical globe, by the relation

$$y = \arg \mathrm{gd}\, x \text{ or } y = \log \tan \left(\frac{x}{2} + \frac{\pi}{4}\right).$$

Thus it is possible to determine the positions of the parallels by calculation.

[1] Hardy, *Pure Mathematics*, p. 377. The meaning is "y is the argument whose Gudermannian is x".

SECTION VIII.

LIMITS.

THE EXERCISES OF SECTION VIII.

*** The numbers in ordinary type refer to the pages in *Exercises in Algebra*, Part II; the numbers in heavy type to the pages of this book.

EXERCISES		PAGES
CXI.	The Meaning of a "Limit"	324, 541
CXII.	Differentiation (I)	338, 549
CXIII.	Differentiation (II)	346, 549
CXIV.	Integration	354, 553
CXV.	Differential Equations	364, 556
CXVI.	Some Theoretical Considerations	376, 557
CXVII.	A General Formula for Expansions	389, 559
CXVIII.	Supplementary Examples—	
	A. Differentiation	403, 560
	B. Integration	405, 560
	C. Differential equations	407, 560
	D. Partial differentiation	413, 563
	E. Total differentials	420, 563
	F. Various practical applications	422, 563

CHAPTER LI.

DIFFERENTIATION AND INTEGRATION.

§ 1. Ex. CXI. *The Nature of a Limit.*—The doctrine of the differential and integral calculus in the proper sense of the term differs from the "calculus of approximations" studied in previous sections by the fact that it is founded upon the conception of a limit. The student has been prepared for this fundamental idea at various points of his earlier work; Ex. CXI is devoted entirely to the important task of making it as definite and clear as possible.

The remark that the real nature of many algebraic ideas has long been obscured for the student by the misapprehensions of the writers who fixed the traditions of the text-book and the classroom is especially true of the idea of a limit. Even now these misapprehensions have not lost their currency. It is particularly important, therefore, that the subject should be represented in the clear outline to which it has been reduced by modern critical mathematicians. Ex. CXI attempts an exposition which shall give the essential features of Cantor's doctrine free from the technical complications which make it a rather formidable affair as it appears in the writings of the more rigorous modern writers.

These essential features—which are really extremely simple—stand out most clearly in instances such as Ex. CXI, Nos. 10, 14, 15. Thus in **No. 10** the numbers obtained by giving to n, in the expression $2 - 1/n$, successively higher integral values form a sequence of rationals which constantly rise in value but have no last term. There is, however, a certain rational number—namely 2—which comes *next after* all possible terms of this sequence. That is to say, if any rational number be named less than 2 there will always be some term of the sequence $2 - 1/n$ between it and the number 2. This is what is meant by calling 2 the *limit* of

the sequence $2 - 1/n$. Similarly in **No. 14** the points to the left of P form an ordered sequence which has no last term towards P, while P is itself the first point of OX which lies beyond all possible terms of the sequence. P is, therefore, the limit of the points which lie on its left. [See ch. xxxix., pp. 414-16.] Again, in **No. 15**, the rationals between 2 and 5 form a sequence which is endless both ways, and the numbers 2 and 5 are themselves the first rationals met with beyond the sequence. These numbers are, therefore, the lower and upper limits of the sequence. The rationals from 3 (inclusive) up to but not including 10 form a sequence which is endless towards 10 but has a definite beginning in 3. Thus it has an upper limit—the number 10 which is the first rational beyond the sequence—but no lower limit. Finally, since the integers between 100 and 200 have both a definite beginning (101) and a definite end (199), they form a sequence which has neither an upper nor a lower limit.

Consideration of these examples enables us to state rather more formally the definition of a limit. (i) A limit is always the limit of a sequence S′ which is thought of as *part of a wider sequence* S. (ii) The sequence S′ must have either no first or no last term. (iii) If S′ has no first term let there be a term L of S which is not a term of S′ but is the last term of S before all possible terms of S′; or if S′ has no last term let there be a term U of S which is not a term of S′ but is the first term of S after all possible terms of S′. Then L and U are respectively the lower and upper limits of S′ in the sequence S.

Two points in this definition require emphasis. The first is that the limits L and U are not themselves members of the sequence S′ but are terms of S which lie outside S′. The neglect of this point is responsible for most of the current inaccuracies in the use of the idea of a limit. The second point follows from the first. The sequence S′ may in itself be capable of having a limit—that is, it may be without a first or a last term or without both—yet the question whether or not it actually has a limit and what that limit is depends upon the sequence S of which it is regarded as forming a part. Both points are well illustrated by Nos. 17, 18. In **No. 17** S′ is the sequence of rationals less than 3 and (since there is no *last* rational less than 3) is obviously capable of

an upper limit. If S, the wider sequence of which it is a part, be taken to be the whole sequence of rationals from zero upwards, then the upper limit is evidently 3, for this is the first term of S beyond all possible terms of S′. But in No. 17 (i) we are told to regard S as a sequence composed of all the rationals less than 3 together with the rationals from 4 upwards. Now 3 is not a member of this sequence, so that it is not the limit of S′. As a matter of fact the first term of the prescribed S which comes after all possible terms of S′ is the rational 4 ; thus 4 is in this case the upper limit of the rationals less than 3. Again in the last part of the same question we are invited to consider the same S′ as part of an S which consists of the rationals less than 3 together with the rationals greater than 4. In this case it is evident not only that neither 3 nor 4 is a term of S but also that there is *no* term of S which can be said to be the first after all possible terms of S′. Thus the rationals less than 3 have, as a section of this new sequence, no limit at all. In **No. 18** the function has already been studied as Ex. LXXV, No. 13 (iii), and its graph is shown as fig. 98 (p. 424 of this book). If we give x values from + 1 to + 1·5 the value of n is, by definition, 2 ; the number under the radical sign will, therefore, be negative and the function has no real values. It has, on the other hand, real values for an endless sequence of values of x between zero and 1. It has again real values for values of x from 2 (inclusive) up to (but not including) 3. It follows that if we regard the numbers which measure the abscissæ as segments of an S which consists of all the real numbers, then the limit of the abscissæ less than 1·5 is 1. On the other hand, if we regard any one of the isolated segments of the graph in fig. 98 as an S′ belonging to an S which is simply the whole of the points which correspond to real values of the function, then it is evident that the limit of the points constituting the first upper segment on the right of the y-axis is the first point in the second segment, that is the point $(+ 2, + \sqrt{2})$. For the first segment (S′) has, as we have seen, no last term, and the first term of S beyond it is the first point of the second segment. This example shows, even better than No. 17, that it is a mistake to suppose that the terms of a sequence necessarily approach "indefinitely near" to their limit, and still more erroneous to say that they "ultimately coincide" with the limit. As we have seen, the

latter statement is never true; for the limit is always outside the sequence of which it is the limit. Whether the former statement is true or not depends upon the nature of the S of which the S′ is conceived as a segment. In most cases of practical importance it *is* true, because the S is either the sequence of real (or else of rational) numbers or some other sequence whose terms have one-to-one correlation with the number-sequence; but it is not *necessarily* true.[1]

It is extremely important to note that in the foregoing discussion S′ and S need by no means be numbers; the definitions and arguments apply equally well to any sequence in which the terms follow one another in accordance with a definite rule of order. The points on a line and the ordinates of a curve will occur to everybody as instances of such non-numerical sequences; it is, however, profitable to quote at least one example outside the ordinary field of mathematical discussion. Consider the notes which a skilled violinist can elicit by bowing (say) the "A" string of his instrument at the same time that he "stops" it at some point with his finger. As he runs his finger up the string these notes will form a sequence S′ regularly ordered in respect of "pitch". They will all be higher in pitch than the note A but they will have no last term towards A. Again, this sequence of notes is a segment of a wider sequence S which consists of all possible notes producible by means of strings of all possible lengths. Finally, the note A is the first term of S below the sequence S′. Thus it is the lower limit of S′ in exactly the same sense as 5 is the lower limit of the sequence of numbers produced by giving to n in the expression $5 + 1/n$ all possible positive integral values from unity upwards in order.

It is also very important to be clear about a complication which appears constantly in the case of a sequence of numbers, and may appear in other cases—for example, sequences of notes. This is the complication dealt with in Ex. CXI, p. 330, and in the footnote. It emerges there in connexion with the problem of determining the height of a certain ordinate of a curve. The point is that while the ordinates of a

[1] The reader who demands further authority for these statements should consult an admirably simple and lucid article on the subject published by Mr. Bertrand Russell in the philosophical journal *Mind* for April, 1908.

curve are all different *lines* the numbers which measure their heights are not necessarily all different *numbers;* since ordinates having the same height may occur in different parts of the curve. Now the foregoing discussion of limits presupposes that the terms of S′ and S are all different from one another; only in that case can we say that the limit L or U is not a member of S′. The question arises, therefore, how this condition is to be satisfied in the case of the numbers which measure ordinates whose heights may recur. To answer it we note that in the case of these numbers the S which we have in mind—the wider sequence to which the limit belongs—is always the whole sequence of real numbers in their natural order. It follows that the S′ must be simply a segment of this sequence. In order, therefore, to apply the notion of a limit we must select from the whole sequence of numbers measuring the heights of the ordinates a segment in which the numbers are all different and all either increasing or decreasing in the natural order; and this segment is to be taken as the S′ to which our reasonings apply. As is shown in Ex. CXI, p. 330, such a segment can always be found however "wavy" the curve may be, provided that the waves have definite dimensions. Difficulty will, in fact, arise in only three cases. (1) A curve may have the infinite waviness of the "crinkly curves" of Ex. CXVI and their congeners. In this case (discussed in the next chapter) there actually is no limit. (2) The numbers may be *all* the same. In this case the notion of a limit as here expounded cannot be applied. On the other hand, as is shown in the footnote, *Exercises*, II, p. 331, we can do perfectly well without it. (3) The terms of the sequence may, as in the case discussed in the Note after No. 12, appear alternately above and below the limit. In this case we may regard them as constituting two sequences in which the upper limit of one coincides with the lower limit of the other.

It is only in the last case that the definitions of a limit found in modern treatises[1] show a clear advantage over the

[1] "A function $f(x)$ has the limit L at a value a of its argument x, when in the neighbourhood of a its values approximate to L within *every* standard of approximation." (Whitehead, *Introduction*, p. 229.) "If a variable x represents any number of a sequence $a_1, a_2, a_3, \ldots, a_n, \ldots$, it is said to approach a number a as a *limit*, provided that, corresponding to every positive number, there

Cantorian definition of § 4. The reader may demur to this remark and point out that the adoption of the usual definition would also make it unnecessary to exclude the case in which all terms of the sequence are identical. For example, the definitions given in the footnote and other recognized equivalents of them would give 2 as the limit of the endless sequence

$$2, 2, 2, 2, 2, \ldots\ldots$$

To this objection it is perhaps sufficient to reply (1) that none of the ways of regarding limits which these definitions imply is comparable in clearness and interest with the Cantorian notion in its direct and simple form; (2) that the gain in generality obtained by adopting them is purchased, as far as the beginner is concerned, at far too great a price, and (3) that they are apt to obscure for the beginner the vital fact that the use of limits in calculations leads to *exact* results, and to leave him with the notion that the calculus after all only gives approximate results, though these may have indefinite closeness to the truth. These arguments are offered as a defence of the policy of making the view of limits presented to the beginner in Ex. CXI the standard definition to be used in all problems that meet him at the present stage of his studies. The usual definition is introduced in the Note on No. 12 mainly to prevent confusion if the student should come across it in his mathematical reading.

It is not superfluous to add—for it is frequently forgotten—that the definitions of the footnote above are specially intended to supply a necessary and sufficient test for a limit in the case of a *function*;[1] that is, in a case where the S is necessarily the whole sequence of real numbers. Thus (as Nos. 17, 18 show) it is not really so general as the definition adopted as our standard, and is prone to give the student too limited a view of the scope of the idea of a limit.

§ 2. Ex. CXI. *The Practical Uses of Limits.*—It was convenient to begin this chapter with a theoretical discussion of the nature of limits. In accordance, however, with our universal pedagogical principle, Ex. CXI itself begins with the dis-

exists a number m such that the numerical value of the difference $a - a_n$ is less than m, provided only that n is greater than m." (Young, *Concepts*, p. 204.)

[1] Russell, *Principles*, p. 327, and references.

cussion of problems intended to prepare the student for the notion of a limit by demonstrating its usefulness. Speaking broadly, that usefulness falls under two heads which may be briefly considered.

(1) The first head is illustrated by the use of a limit to define the "velocity of a moving point at a given moment". The discussion of Ex. CXI shows that, if we define velocity as the quotient of a distance travelled by the time in which it is traversed, then the "velocity at a given moment" is not a velocity at all. On the other hand, if we consider the distances travelled by the point during a series of constantly decreasing intervals of time and divide each distance by the length of the corresponding interval, we shall again fail, as a rule, to obtain anything that can be called "the" velocity of the point, for all the results will be different except in the special case of "uniform" motion. But if the sequence of "average velocities" thus calculated follows some definite law of succession as the interval is taken smaller, then it will generally have a definite limit as the interval approaches zero. Thus the limit is a perfectly definite number associated in a perfectly unambiguous way both with the given moment and with the endless sequence of different average velocities. Moreover, for small intervals of time the average velocities are sensibly equal to the limit, the differences being of theoretical rather than of practical importance. It follows that, although the "velocity at the given moment" is not really a velocity at all, it is quite the most useful number to quote in order to describe the behaviour of the moving point while it is in the neighbourhood of the place which it occupies at the given moment. A similar statement explains the practical value of the derivative or "differential coefficient" of a function in other cases. If, for instance, the value of x in the function $\sin x$ is increased to $x + h$ the ratio $\{\sin (x + h) - \sin x\}/h$ is quite ambiguous in value; for it depends upon the value of h. But as h approaches zero the ratio has a limit, $\cos x$, and although this limit is not any one of the ratios yet it is connected in an unambiguous and unique way with the sequence of actual ratios. It may, therefore, be regarded as representing them just as the so-called "velocity at a given moment" represents the endless sequence of average velocities during intervals of time which succeed that moment.

(2) The second practical use of limits comes into view when we employ them to determine a magnitude which cannot be evaluated directly. We have already had important instances of this kind of calculation, the most striking, perhaps, being the calculation of the area of a spherical belt in Ex. LXXXVI. The method was analysed on p. 444 of this book, and it is unnecessary to repeat the analysis here; the teacher should, however, make sure that he appreciates the argument as set forth in the first two paragraphs of that page. It is necessary to add only (1) that in most cases one of the two sequences is the sequence of real numbers, the other being a sequence containing the magnitude to be determined; (2) that the indirect calculation is most convincing, especially to the beginner, if the magnitude (M) and the number (N) which measures it can be exhibited, like the S and C of p. 444, as filling corresponding gaps in sequences which extend on both sides of them—that is, if M and N can be exhibited as at once upper and lower limits of sequences known to correspond to one another term by term; (3) but that the argument is unchanged in principle, though less striking, if, through necessity or for brevity's sake, M and N are regarded as the limits (upper or lower) of sequences which lie only on one side of them. As an illustration of the last remark consider the example in Ex. CXI, § 2. Here M, the magnitude to be determined, is the height of PQ; N is the number which measures it. It is shown (1) that the ordinate whose magnitude is in question lies between, and is the limit of, a lower sequence consisting of ordinates pq and an upper sequence consisting of ordinates $p'q'$, (2) that N lies similarly between, and is the limit of, the sequences of numbers represented respectively by

$$\{3x^2 - h(3x - h)\}a \text{ and } \{3x^2 + h(3x + h)\}a$$

and (3) that the latter sequences correspond to the former, term by term. From these premises it is a very convincing deduction that the height of PQ is exactly $3ax^2$; for PQ is the only line between the two former sequences and $3ax^2$ the *only* number between the two latter. On the other hand, the argument would have been perfectly sound if we had contented ourselves with pointing out that the ordinates $p'q'$ and the numbers $\{3x^2 + h(3x + h)\}a$ are sequences corresponding term by term, that PQ is the limit of the former sequence and $3ax^2$ of the latter, and that therefore they correspond.

As is pointed out in Ex. CXI at the end of § 4 this briefer argument is always sufficient unless the sequence to which M belongs is discontinuous so that there are, for certain values of the variable, two limits corresponding to the two ways in which that value may be approached.

§ 3. Exs. CXII, CXIII. *Differentiation.*—In the next two exercises the preceding theory of limits is applied to the technical problems of determining the derived functions or derivatives of functions of given form. The teacher will note that the idea of the derived function is introduced to the student as a generalization of the familiar ideas of connexions between area-functions and ordinate-functions, ordinate-functions and gradient-functions, etc. No special notations have hitherto been employed to symbolize these different notions; the need of them has not been felt and their absence has probably been an aid rather than a hindrance to clear thinking. But with the generalization of these special functional relations into the idea of a derivative a specific notation becomes necessary. The practice of mathematicians offers a choice from several alternative notations of which $D(y)$, y' and dy/dx are the most important. The student is to be taught eventually to use each of these forms of symbolism, but the first is introduced as the standard form for reasons, positive and negative, that must be briefly indicated.

The negative reasons may be considered first. The notation dy/dx, which undoubtedly has played and continues to play the most important part in mathematical literature, goes back, as is well known, to the mathematician and philosopher Leibniz, who shares with his contemporary Newton the credit of having invented the differential calculus as a distinct branch of mathematics. It is not so well understood that it expresses a view of the nature of a "differential coefficient" which is quite out of harmony with modern ideas and, in particular, conflicts with the doctrine of limits explained in §§ 1, 2. Briefly, the view was that any finite value of the variables y and x is really the sum of a vast number of "infinitesimal" values which, though immeasurably small, have yet a definite magnitude—much as (to use an illustration given by Leibniz in another context) the sound of the sea is the sum of a vast number of sounds which, though individually inaudible, must really exist and have a definite degree of loudness. It was assumed that the

"infinitesimals" of a given variable all have (like the "infinitesimal" atoms of a given substance) the same magnitude. Thus the differential coefficient dy/dx is simply the ratio of the "infinitesimals" of the two variables in the case in question—the ratio being finite and measurable, just as the relative weights of atoms are measurable, in spite of the smallness of the terms. It need hardly be said that this view is no longer held.[1] The student of the differential calculus is always warned that dy and dx in the expression dy/dx are *not* to be regarded as "infinitely small" numbers, or, indeed, as numbers at all, and that dy/dx is not a ratio but only the *limit* which the ratio of the increments of the variables approaches as the increment of x approaches zero. But in spite of this warning the erroneous presuppositions of the notation still produce a confusing effect upon the student's mind. From the pedagogical point of view, to make the student write the derivative as a fraction and at the same time to forbid him to think of it as a fraction is a poor plan.

Moreover, any virtue which it may have is effectually destroyed if, in spite of the protest, the student *is* taught to treat dy/dx as a fraction in arguments of the following type :—

$$\begin{aligned} dy/dx &= 3x^2 \\ dy &= 3x^2 \,.\, dx \\ y &= x^3 + \mathrm{C}. \end{aligned}$$

It is scarcely a cause for wonder that few elementary students have really clear ideas as to how the calculus "works," even though familiarity with its applications may have convinced them that in some mysterious way it produces trustworthy results. There is only one way of avoiding this unsatisfactory state of affairs—namely, to avoid the notation which creates it.

For these reasons the Leibnizian notation is to be withheld until the student's grip of the logic of the calculus is strong enough to withstand its misleading suggestions. For the first stages of his progress we must use either $D(y)$ or y' as our symbol of differentiation. The second symbol has the advantage of compactness but is not nearly so expressive as the former. The notation $D(y)$ constantly holds before the student's mind the fact that the object of inquiry is a *function*

[1] For criticisms of it see Russell, *Principles*, chs. XXXIX.-XLI., or the more popular account in Whitehead, *Introduction*, pp. 224-7.

which he is to derive from the given *function y* by means of a definite rule of procedure. Since this relationship between *functions* is the essence of the whole matter (see p. 248 of this volume), the notation which so directly suggests it is incomparably the best for the beginner. Moreover, by the simple device of inverting the D we have (as in Ex. CXIV) a symbol which suggests in the clearest manner the fact that in the process of integration we are merely tracing in the reverse direction the relation between a function and its derivative.

We turn now to the technical rules for differentiating a given function. It should be noted that the student is already acquainted with the *results* of differentiating most of the standard forms. His business here is, first, to use the doctrine of limits to place those results upon a proper logical basis and, second, to apply them to the ready determination of derivatives in the more complicated cases. For both purposes it is necessary to establish certain simple but most important theorems about the sum, product, and quotient of limits. Consider the problem of finding the derivative of $\sin x$ as it is treated in the older textbooks. The argument runs as follows :—

$$\begin{aligned}\frac{dy}{dx} &= \operatorname*{Lt}_{h\to 0} \frac{\sin(x+h) - \sin x}{h} \\ &= \operatorname*{Lt}_{h\to 0} \cos(x + h/2) \cdot \frac{\sin(h/2)}{h/2} \\ &= \cos x.\end{aligned}$$

The important point is the step from the second to the third line. The argument is that since the limit of $\cos(x + h/2)$ is $\cos x$ and the limit of $\sin(h/2)/h/2$ is unity the limit of their product is the product of $\cos x$ and unity. But it is clear that this conclusion only follows if the limit of the product of two factors is identical with the product of the limits of the factors. No process of differentiation which implies the truth of this proposition can be logically satisfactory until the proposition itself has been proved. In Ex. CXII, **No. 7**, the student is, therefore, called upon to prove it by the method already applied in § 2 of the exercise to the simpler case of the sum of two functions. The proof intended runs as follows :—

Any term of U can be expressed as $L_u + p$, and the corre-

sponding term of V as $L_v + q$. Hence the corresponding term of W may be written

$$L_u \cdot L_v + qL_u + pL_v + pq.$$

By hypothesis p and q represent numbers which both approach zero in a sequence which has no last term. The corresponding terms of W are therefore also a sequence with no last term, for L_u and L_v are constant numbers. If we put $p = q$ = zero the corresponding term of W is $L_u \cdot L_v$, and is obviously the first term of W beyond the sequence just referred to. That is, it is the limit of that sequence. We conclude that as the terms of U and V approach respectively the limits L_u and L_v the terms of W approach as their limit the product $L_u \cdot L_v$.

This fundamental theorem about limits is applied in Ex. CXII, A, to establish some of the simpler standard forms. In Ex. CXII, B, it is used to justify the rule for differentiating a function of a function. In Ex. CXIII it is once more used to prove the familiar rules for differentiating the product and the quotient of two functions. All these matters are of well-recognized importance, and it is necessary only to emphasize the necessity of securing that the logic involved in them is clearly understood.

In Ex. CXIII, B, we return to Wallis's Law and prove it for all rational exponents by a method independent both of the binomial theorem (of which no general proof has yet been given) and of the exponential curve. The new proof is based upon an idea borrowed from Wallis's *Arithmetica Infinitorum*. It is not purely algebraic but involves assumptions much less serious than the assumption that the exponential curve has at every point a definite tangent corresponding to a definite limit of $(a^{x+h} - a^x)/h$. [See Ex. CXVI, § 3.] The proof is prepared for by three examples (Nos. 7, 8, 11). In **No. 7** we assume $D(x^m) = mx^{m-1}$ and $D(x) = 1$. Hence by the rule for differentiating a product, we have

$$\begin{aligned} D(x^{m+1}) &= D(x^m \cdot x) \\ &= x^m \cdot D(x) + x \cdot D(x^m) \\ &= x^m + mx^m \\ &= (m + 1)x^m. \end{aligned}$$

Similarly in **No. 8** we have

$$\begin{aligned} D(x^{p+q}) &= D(x^p \cdot x^q) \\ &= x^p \cdot D(x^q) + x^q \cdot D(x^p) \\ &= x^p \cdot qx^{q-1} + x^q \cdot px^{p-1} \\ &= (p + q)x^{p+q-1}. \end{aligned}$$

No. 11 is established by a similar application of the rule for differentiating a quotient.

§ 4. Ex. CXIV. *Integration.*—Current ideas about the nature of an integral show, like those about the nature of a differential coefficient, traces of the erroneous mathematical philosophy of earlier days. Under the influence of Leibniz the problems first systematically studied by Wallis came to be regarded as having as their aim the summation of an "infinite" number of "infinitesimals" dy of the form $y \,.\, dx$, where dx symbolizes as before the *minima indivisibilia* of which any finite value x of the independent variable is composed. This view is still represented not only by the usual notation

$$\mathrm{I} = \int y \,.\, dx$$

which (like dy/dx) was first introduced by Leibniz, but also by the common statement that an integral is the sum of an infinite number of infinitely small magnitudes. With the rejection of the notion of an infinitesimal as a definite atomic magnitude this statement and the notation which expresses it have both become inadmissible and should certainly be abandoned. If dx has any numerical significance at all it stands for the increment h when h is zero. It follows that the product $y \,.\, dx$ is also zero for all values of y, and, therefore, that the problem represented by

$$\int y \,.\, dx$$

is the summation of a series of zeros. To teach that anything but zero can result from this process, even if the terms are "infinite" in number, is simply to darken counsel.

Expressed in terms of sounder modern ideas the matter may be stated as follows. Let y be any function of x and let δx stand (as in the earlier sections of this work) for a "small" constant increment of the independent variable. Further, let the small product $y \,.\, \delta x$ be the increment of another function I (the integral). Let x assume in succession values from zero up to a final value $x - \delta x$ at intervals of δx, and let y in the products symbolized by $y \,.\, \delta x$ assume in succession the values corresponding to these. Then for every assumed value of the interval δx the sum of the products will have a certain value. As the interval is shortened and approaches zero this sum will (it is assumed) approach a limit I. Then I may be defined as the *integral* of the function y or of the product $y \,.\, \delta x$. Thus I is not the sum

of an infinite number of products; it is simply the *limit* of the sum of a *finite* number of products. Indeed there is neither need nor warrant for introducing the term "infinite" at any point of the definition or discussion. Substituting δx for the nonsensical dx we may still usefully retain the Leibnizian mode of expression—

$$\mathrm{I} = \int y \,.\, \delta x$$

but the symbol "$\int$" must now be read: "limit of the sum as δx approaches zero".

As is well known there is a very striking and simple connexion between the functions I and y: the function y is simply the first derivative of I. Thus there are two alternative ways of regarding an integral of a function y. We may think of it either as the limit of the sum of products of the form $y \,.\, \delta x$ or as the function from which y would be obtained by the process of differentiation. Of these the first way is undoubtedly the more important from the point of view of practical applications; in physics, in mensuration, etc., the integral almost always appears as the limit of a sum. But from the theoretical point of view it is much simpler to regard the relation between a function and its integral as the inverse of the relation between a function and its derivative. The student of this book has been familiar from the middle of Part I onwards with the ideas which find their logical completion in the former, more complex notion. There need, therefore, be no hesitation in choosing at this stage the better theoretical mode of approach. Thus in Ex. CXIV the integral is first treated as the inverse of the derivative; the demonstration that it may also be treated as the limit of a certain sum is reserved till the end of the exercise.

A remark has already been made about the notation employed in the earlier part of the exercise. When writers on formal logic employ a certain letter to symbolize a given relation, they not infrequently invert the letter to symbolize the inverse relation. That device is followed here. The fact that a function I is the integral of a function y is symbolized by the notation

$$\mathrm{I} = \mathrm{D}(y).$$

This use of the inverted D is believed to be an innovation. The author also believes that it will be found to justify itself by the directness with which it suggests to the student the nature of the problem he is called upon to solve.

It is probably unnecessary to direct attention to any of the examples, which are all limited to results common to elementary treatises on the calculus. It may, however, be pointed out that the form of argument used in division C to show that the integral may also be regarded as the limit of a sum has been chosen chiefly because it seems a natural completion of the arguments of Wallis with which the student's earlier work may be assumed to have made him familiar. In § 3 a special case is treated as an illustration of the general principle. Further concrete illustrations are suggested in **No. 21** and **Nos. 24-26**. The generalization is required in **No. 30**. The proof that

$$f(x) = \operatorname*{Lt}_{h \to 0} \{F(x + h) - F(x)\}/h$$

demands, of course, nothing but the repetition of the argument of Ex. CXI, § 2, in a generalized form. The second part of the argument may run as follows:—

Select any value of h which is an exact submultiple of $x - a$, and erect between the points on the x-axis where $x = a$ and where x has its final value x a series of rectangles of width h, the heights being in succession $f(a)$, $f(a + h)$, $f(a + 2h)$, . . . , $f(x - h)$. Across these rectangles draw a smooth curve in such a way that the area under the curve above any of the segments h of the base is equal to the rectangle standing on that segment. Then it is clear (1) that, if x be now taken to mean the abscissa of the left-hand side of any rectangle, the curve will cut the top of that rectangle and so have $f(x)$ for its ordinate at some point whose abscissa lies between x and $x + h$; (2) that as h approaches zero the limit of this state of affairs is one in which $f(x)$ is the ordinate where the abscissa is x for *all* values of x between the ends of the curve. It is also evident (3) that the total area under the curve is, for all values of h,

$$\operatorname*{Lt}_{h \to 0} \sum_{x=a}^{x=x-h} f(x) \cdot h$$

and (4) that as h approaches zero this sum has as its limit the area under the curve whose ordinate-function is $f(x)$. But by hypothesis that area is $F(x) - F(a)$. It follows, therefore, that

$$\operatorname*{Lt}_{h \to 0} \sum_{x=a}^{x=x-h} f(x) \cdot h = F(x) - F(a).$$

Now, by definition, the relation

$$f(x) = \operatorname*{Lt}_{h \to 0} \{F(x + h) - F(x)\}/h$$

may be expressed by the notation

$$F(x) = \mathcal{C}f(x)$$

and the relation

$$\operatorname*{Lt}_{h \to 0} \sum_{x=a}^{x=x-h} f(x) \,.\, h = F(x) - F(a)$$

by the notation

$$\int_a^x f(x) \,.\, \delta x = F(x) - F(a).$$

We conclude, then, that in all cases the functions intended by the notations $\mathcal{C}f(a)$ and $\int f(x) \,.\, \delta x$ are identical.[1]

§ 5. Ex. CXV. *Differential Equations.*—Under the heading of "differential formulæ" the student has already faced the fundamental questions involved in the solution of differential equations. He should be able, therefore, without much difficulty, to acquire the modest amount of technique demanded by the examples of Ex. CXV. These are limited almost entirely to differential equations which play a part of fundamental importance in the theories of mechanics, physics, and engineering.

The treatises on the *Differential and Integral Calculus* by Granville (Ginn & Co.) and by Hulburt (Longmans) both contain chapters dealing with the subject in a simple way. The teacher who desires more information than is given in the text or in ch. LII., § 4, will probably find what he wants in either of them. Prof. Perry's well-known *Calculus for Engineers* (Macmillan) contains a number of interesting applications rather too technical in character for inclusion here.

[1] The reader may notice an imperfection in this argument: namely, the assumption that the ordinates are equidistant. For complete rigour it should be shown that $\Sigma f(x) \,.\, h$ tends to the same limit however the ordinates are distributed.

CHAPTER LII.

EXPANSIONS. SUPPLEMENTARY EXAMPLES.

§ 1. Ex. CXVI. *Some Theoretical Considerations.*—Expansions, regarded as approximation-formulæ, have engaged our attention from time to time since the earliest exercises of the course. The final substantive exercise in this section is to be devoted to the quest for a generalization that shall include all the individual expansions in one comprehensive formula and shall, at the same time, provide an universal test of their usefulness and validity. This imperial formula is, of course, Taylor's Theorem.

Taylor's Theorem is best regarded as a deduction from, or an application of, the Theorem of Mean Value. The first division of Ex. CXVI is given, therefore, to the study of that familiar proposition, including the necessary inquiry into the conditions under which it holds good. Here is a favourable opportunity for introducing the student to a topic which illustrates, better almost than any other, the significance of the "rigorist's" suspicion of proofs based upon "intuition" —illustrates it by an instance in which it is strikingly justified. This topic is the fascinating one of tangentless curves. "If we draw a curve with one stroke of a pencil from A to B so that for one value of x there is only one value of y, and there is no kink or sudden bend in the curve, we can easily satisfy ourselves that there is always one point at least on the curve the tangent at which is parallel to AB." Yet when the student has made acquaintance with the cases, produced by Weierstrass and others, in which this seductive conclusion is demonstrably wrong, he will be readier to accept the rigorist's contention that "you cannot prove anything by an appeal to the eye" and to understand why he "tries to live up to the sentence 'point d'images dans cet ouvrage' and to state precisely the conditions that are vaguely implied in the

words 'drawn with one stroke of the pencil,' 'no kinks or sudden bends,' and so on ".[1]

The question of tangentless curves is treated too fully in the text to require commentary. Weierstrass's epoch-making paper, "Ueber continuirliche Functionem eines reelles Arguments die für keinen Werth des letzteren bestimmten Differentialquotienten besitzen," is reprinted in his collected *Werke*, Vol. II. The argument is neither abstruse nor very long, but cannot be summarized with enough brevity for reproduction. Probably the graphic analysis given in *Exercises*, II, figs. 112-13, will be sufficiently illuminating.

The account of the "crinkly curves" of Moore and the "space-filling curve" of Peano is taken from an interesting paper by the former in the *Trans. Amer. Math. Soc.*, Vol. I, 1900. It will be noted that Moore's two curves may be regarded as giving respectively the ordinates and abscissæ of Peano's. Fig. 140 shows one-ninth of the abscissa-curve, $x = \phi(t)$, at the stage described in **No. 16**. Fig. 141 is the diagram asked for in **No. 20**, the abscissæ of successive nodes being the *ordinates* of the nodes in fig. 140 and the ordinates being the ordinates of the nodes in *Exercises*, II, fig. 115. The figure is, therefore, an early stage in the development of Peano's curve towards the limit in which it passes through every point of the square. Since every point on Peano's curve is given by a distinct value of t and every value of t gives a point on the curve, it is clear that the total number of points in the square is the same as the number of possible values of t. But each value of t corresponds to a single point of the line OT; hence the number of points in the square is the same as the number of points in the line. Cantor showed, by a different line of argument, that this result may be extended to the number of points in a cube.

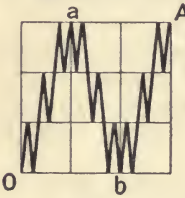


FIG. 140.

Gosiewski's line composed of "condensed semi-circles" always proves a very fascinating topic of discussion. The author owes his knowledge of it to Dr. L. Silberstein.

[1] The words placed between inverted commas are quoted from a manuscript note kindly placed at the author's disposal by Mr. C. S. Jackson.

§ 2. Ex. CXVII. *A General Formula for Expansions.*— As is well known, Brook Taylor, in his *Methodus Incrementorum Directa et Inversa* (1715) gave his expansion without any formula for estimating the remainder. It is still usual for text-books to begin by deriving the coefficients on the assumption that the expansion is valid, and to keep the question of the remainder for subsequent discussion on the lines set by Lagrange or Cauchy. The proof given in the text aims at developing at once the expansion with its remainder-formula. The chief reason for adopting this plan

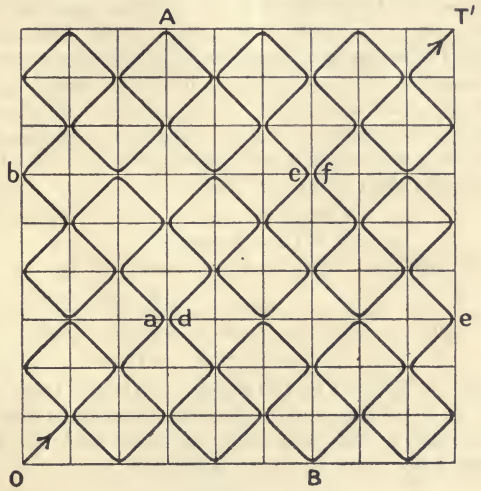


Fig. 141.

is that, at the stage which the student has now reached, the value of the theorem consists no more in its giving a general expansion-formula than in supplying a general test of the validity of such expansions. Moreover, it will probably be found that the argument given, though it appears long, carries more conviction to the student's mind than the ordinary proof, because it gives a fuller insight into the reasons why the formula should have such wide applicability. It will be seen that it is based upon the theorem of Wallis of which so much use has already been made and that it deals with the remainder at each of the stages into which the derivation of the whole expansion-formula is divided.

The usual method of deriving the theorem is given as an exercise for the student in **Nos. 24-6.** The teacher who prefers that plan may easily turn first to those examples and consider the fuller proof later. In that case it would be advisable to let the class use the remainder-formula without proof in Exs. 1-23.

§ 3. Ex. CXVIII. *Supplementary Examples*, A, B.—In division A of the supplementary examples the student is introduced for the first time to the ordinary, or Leibnizian notation for differentiation. The first application, in **Nos. 1-5,** is to the differentiation of an inverse function—a case in which its usefulness is apparent. In **Nos. 6-7** the hyperbolic functions are considered for the first time in this section, though their differential formulæ were, it will be remembered, deduced in Ex. CIX, p. 303. The method of logarithmic differentiation is also explained. In **No. 10** (vi), (vii), for the functions gd x and arg gd x see Ex. CX, D. The solution of **No. 10** (viii) is as follows :—

$$\begin{aligned} y &= 2 \text{ arc cos } \sqrt{[(x-b)/(a-b)]} \\ x &= b + (a-b)\cos^2\tfrac{1}{2}y \\ dx/dy &= -2(a-b)\cos\tfrac{1}{2}y \,.\, \tfrac{1}{2}\sin\tfrac{1}{2}y \\ &= -(a-b)\,.\,\frac{(x-b)^{\frac{1}{2}}}{(a-b)^{\frac{1}{2}}}\cdot\frac{(a-x)^{\frac{1}{2}}}{(a-b)^{\frac{1}{2}}} \\ &= -\sqrt{\{(a-x)(x-b)\}}. \end{aligned}$$

Since $dy/dx = \mathrm{L}t\delta y/\delta x = 1/(\mathrm{L}t\delta x/\delta y) = 1(dx/dy)$ the required derivative is the reciprocal of the present result. **No. 10** (ix) is done similarly, with the exception that sinh $\frac{1}{2}y$ is to be determined from cosh $\frac{1}{2}y$ by the relation

$$\cosh^2\tfrac{1}{2}y - \sinh^2\tfrac{1}{2}y = 1.$$

The novelty in division B is the integration of a function of a function, usually called "integration by substitution". The name here used has the advantage of bringing out the relation of the process to that of differentiating a function of a function. The examples cover the more important forms studied in an elementary course and should be done very thoroughly, integrations found difficult at the first attempt being written out until mastered.

§ 4. Ex. CXVIII, C. *Differential Equations.*—Division C supplements the examples of Ex. CXV in one or two important respects. In **Nos. 21-5** the student learns the general method of solving an equation of the first degree

with constant coefficients by the substitution $y = e^{px}$. The teacher should refer here to Ex. CXV, Nos. 24-6, and show how the use of the exponential expressions for the sine and cosine makes possible a remarkable simplification of treatment. In **No. 23** this principle is to be combined with that of Ex. CXV, § 3. For instance in (iii), to find the particular integral we assume $y = a \cosh \frac{3}{4}x + b \sinh \frac{3}{4}x$, and obtain

$$\{4(\tfrac{3}{4})^2a - \tfrac{3}{4}b - 5a\}\cosh \tfrac{3}{4}x + \{4(\tfrac{3}{4})^2b - \tfrac{3}{4}a - 5b\} \sinh \tfrac{3}{4}x = 2 \cosh \tfrac{3}{4}x$$

whence we derive the relations

$$4(\tfrac{3}{4})^2a - \tfrac{3}{4}b - 5a = 2 \text{ and } 4(\tfrac{3}{4})^2b - \tfrac{3}{4}a - 5b = 0$$

giving $a = -11/14$ and $b = +3/14$. So far the particular integral. For the complementary function we assume that $y = e^{px}$ satisfies

$$4y'' - y' - 5y = 0$$

and deduce the condition

$$4p^2 - p - 5 = 0$$

whence $p = 5/4$ or $p = -1$. Thus the full solution of the equation is

$$y = Ae^{5x/4} + Be^{-x} - \frac{11}{14} \cosh \tfrac{3}{4}x + \frac{3}{14} \sinh \tfrac{3}{4}x$$

A and B being arbitrary constants.

In **No. 25** (iii) we must assume for the particular integral, in accordance with Ex. CXV, No. 31,

$$y = ax + b + ce^{-5x}.$$

Substitution in the equation gives

$$121ce^{-5x} + ax - 4a + b = 2x + e^{-5x}$$

whence $c = 1/121$, $a = +2$ and $b = +8$. For the complementary function we assume $y = e^{px}$ and obtain, on substitution, the quadratic

$$4p^2 - 4p + 1 = 0$$

whose roots are both $+\frac{1}{2}$. Hence, in accordance with **No. 24**, the full solution is $y = (A + Bx)e^{\frac{1}{2}x} + e^{-5x}/121 + 2x + 8$.

In § 3 we attack the case, so important in physical and other applications, where it is necessary to "separate the variables". It should be noted that the customary practice of separating the dy and dx is quite indefensible except upon the understanding that, when separated, they are really used instead of δy and δx. It seems to the interest of clear thinking that at first, at any rate, the student should actually make the substitution of the differentials.

No. 26 (vi)-(x) are, it will be noticed, all of a single type which may be represented by the general equation

$$dy/dx = y(x + a)/(b_0 + b_1x + b_2x^2).$$

Written in this form it plays an important part in Prof. Karl Pearson's theory of "skew frequency-curves" to be considered in Section IX. The method to be followed is essentially the same in each case. Thus in (vii) we have, after separating the variables,

$$\frac{\delta y}{y} = \frac{2 \,.\, \delta x}{x} - \frac{3 \,.\, \delta x}{x^2}$$

whence $\log y = C + 2 \log x + 3/x$

or $\log y = \log A + \log x^2 + \log e^{3/x}.$

So that

$$y = Ax^2 . e^{3/x}.$$

Similarly in No. 26 (x) we have

$$\frac{\delta y}{y} = \frac{(x + 4) \,.\, \delta x}{(x - 1)^2 + 5^2}$$

$$= \frac{(x - 1) \,.\, \delta x}{(x - 1)^2 + 5^2} + \frac{5 \,.\, \delta x}{(x - 1)^2 + 5^2}$$

whence, by **Nos. 12, 14**,

$$\log y = C + \tfrac{1}{2} \log (x^2 - 2x + 26) + \text{arc tan } [(x - 1)/5].$$

So that, finally, $y = A(x^2 - 2x + 26)^{\frac{1}{2}} . e^{\text{arc tan } [(x-1)/5]}.$

§ 4 gives a few simple instances of a process of much mathematical interest and of real importance to the student of physics. From the purely mathematical point of view their interest consists in the demonstration that a new function may be defined, and its properties explored, by means of a differential equation, even in the case when the function is not expressible in terms of functions already known but only as an endless power-series. The Bessel function is one of the most important functions of this kind, and the one which the ordinary student is most likely to meet in physics.[1] From the definition of a simple case of the function in **No. 29** we assume

$$y = c_0 + c_1x + c_2x^2 + c_3x^3 + \ldots$$

and deduce that

$$x(2c_2 + 3 \cdot 2c_3x + \ldots) + (c_1 + 2c_2x + 3c_3x^2 + \ldots) + p^2(c_0 + c_1x + c_2x^2 + c_3x^3 + \ldots) = 0.$$

[1] See Rayleigh, *Theory of Sound*, i., ch. IX., or Gray and Mathews, *Treatise on Bessel Functions*.

From this relation we have $c_1 + p^2 c_0 = 0$, etc., as stated, and all the other coefficients become expressible in terms of c_0.

§ 5. Ex. CXVIII, D, E. *Partial Differentiation, Total Differentials.*—Partial differentiation has been anticipated in several of the examples of Ex. CX, A, B. The present examples put the topic upon its proper basis, and serve, in particular, to extend the applications of the earlier exercises to functions of more than one variable. For this reason alone it would be well worth while giving to the subject the modest amount of time it demands. Much higher claims can, however, be made for it; the student who has no acquaintance with partial differentiation is still too slightly equipped for many important and interesting excursions in elementary physics. The subject is treated fully in some of its more fundamental aspects and is not likely to cause difficulty. Special attention should be given (i) to the conditions for a turning value in the case of a function of two variables—namely, that the partial derivatives of the function with respect to *both* variables must vanish, and (ii) to the method of finding the derivative of an implicit function (**No. 40**).

The total differential is the natural complement of the partial derivative and is, therefore, usefully studied beside it. Speaking generally, the subject has not the wide importance of partial differentiation, but the student of physics has to have clear elementary views about it. Perhaps the treatment given in division D is the most suitable for the purpose. In most physical applications the magnitude of some entity, for example potential energy, is a function only of the coordinates of the points at which its value is considered. In that case its values mark out surfaces in space over which they are constant and the ideas of division D are directly applicable.

§ 6. Ex. CXVIII, F. *Fourier's Theorem.*—This has already been treated in a simple way in Ex. CVI and ch. XLVIII. of this book. The novelty is the solution of simple problems of harmonic analysis by an algebraic instead of a graphic process. It should be understood that such solutions are possible only if the form to be analysed is expressed as a definite function. In most of the practical applications of Fourier's Theorem, for example in tidal prediction, this condition is not fulfilled. In such cases the method of Ex. CVI is, in principle, the one that has to be adopted.

The solution of **No. 47** includes that of **No. 46**. For the first part of the string we have $y = ax/b$ and for the second part $y = a(l - x)/(l - b)$, while the assumption is that the whole shape of the string is given by a sum which may be written for brevity

$$y = \Sigma c_p \cdot \sin \frac{p\pi x}{l}.$$

To determine the value of a specified coefficient c_s we are to suppose a solid to be formed of which the elevation is the triangle formed by the string and the base the curve $y = \sin \frac{s\pi x}{l}$. We then have as one expression for the volume of the solid :—

$$\begin{aligned}
V &= \int_0^b \left[\frac{ax}{b} \cdot \sin \frac{s\pi x}{l}\right] + \int_b^l \left[\frac{a(l-x)}{l-b} \cdot \sin \frac{s\pi x}{l}\right] \\
&= \frac{a}{b} \cdot \int_0^b \left[x \cdot \sin \frac{s\pi x}{l}\right] + \frac{al}{l-b} \cdot \int_b^l \cdot \left[\sin \frac{s\pi x}{l}\right] \\
&\qquad - \frac{a}{l-b} \cdot \int_b^l \left[x \cdot \sin \frac{s\pi x}{l}\right] \\
&= \frac{a}{b} \cdot \left[\left(\frac{l}{s\pi}\right)^2 \cdot \sin \frac{s\pi x}{l} - \left(\frac{l}{s\pi}\right) x \cdot \cos \frac{s\pi x}{l}\right]_0^b \\
&\qquad - \frac{al}{l-b} \cdot \left[\frac{l}{s\pi} \cdot \cos \frac{s\pi x}{l}\right]_b^l \\
&\qquad - \frac{a}{l-b} \cdot \left[\left(\frac{l}{s\pi}\right)^2 \cdot \sin \frac{s\pi x}{l} - \left(\frac{l}{s\pi}\right) \cdot x \cos \frac{s\pi x}{l}\right]_b^l \\
&= \left(\frac{l}{s\pi}\right)^2 \cdot \sin \frac{s\pi b}{l}\left[\frac{a}{b} + \frac{a}{l-b}\right] + \left(\frac{l}{s\pi}\right) \cdot \cos \frac{s\pi b}{l}\left[-a\right. \\
&\qquad \left. + \frac{al}{l-b} - \frac{ab}{l-b}\right] \\
&\qquad - \left(\frac{l}{s\pi}\right) \cdot \cos s\pi \left[\frac{al}{l-b} - \frac{al}{l-b}\right] - \frac{a}{l-b} \cdot \left(\frac{l}{s\pi}\right)^2 \sin s\pi \\
&= \frac{al \cdot l^2}{(s\pi)^2 \cdot b(l-b)} \cdot \sin \frac{s\pi b}{l}
\end{aligned}$$

for the second and third terms vanish algebraically and the last term vanishes because $\sin s\pi = 0$ whenever s is integral.

But we have another expression for this same volume, namely

$$V = \int_0^l \Sigma c_p \cdot \sin\frac{p\pi x}{l} \cdot \sin\frac{s\pi x}{l} = \tfrac{1}{2}\int_0^l \Sigma c_p \left[\cos(p-s)\frac{\pi}{l} - \cos(p+s)\frac{\pi x}{l}\right].$$

Now unless $p = s$ the integrals of all the cosine-functions vanish. The expression for the volume reduces, therefore, as in Ex. CVI to

$$V = \tfrac{1}{2}\int_0^l c_s \cdot \left(1 - \cos\frac{2s\pi x}{l}\right)$$
$$= \tfrac{1}{2}lc_s.$$

Equating the two expressions for the volume we have

$$\tfrac{1}{2}lc_s = \frac{al \cdot l^2}{(s\pi)^2 \cdot b(l-b)} \cdot \sin\frac{s\pi b}{l}$$

or

$$c_s = \frac{2al^2}{(s\pi)^2 \cdot b \cdot (l-b)} \cdot \sin\frac{s\pi b}{l}.$$

The case when $l = 60$, $a = 2$, $b = 24$ was worked out by our simpler method in Ex. CVI and the solution there obtained is shown in fig. 135 of this book (p. 525).

In **No. 48** we must make the most general assumption, i.e.

$$\frac{x^2}{2a} = a_0 + a_1 \sin x + a_2 \sin 2x + a_3 \sin 3x + \ldots$$
$$+ b_1 \cos x + b_2 \cos 2x + b_3 \cos 3x + \ldots$$

To find a_0 we integrate the equivalence, as it stands, from 0 to 2π and obtain

$$\frac{1}{2a}\int_0^{2\pi}(x^2) = 2\pi a_0$$

$$a_0 = \frac{1}{2a} \cdot \frac{4\pi^2}{3}.$$

To find a_s we multiply by $\sin sx$; to find b_s by $\cos sx$. In this way the expansion given in the question is obtained without difficulty.

§ 7. Ex. CXVIII, F. *Curvature.*—This, like other topics in the present exercise, is introduced mainly on account of its importance in applied mathematics and physics. **Nos. 56-7** are, however, set as an interesting addendum to division D. **Nos. 58-60** are very important since they extend to a point moving along *any* curve the property demonstrated in Ex. CXV, Nos. 15-18, only for the case of the circle. In

Exercises, II, fig. 121, take as y-axis the momentary normal at R, and for x-axis any line at right angles to it. Then the equivalence of **No. 58** follows by putting $dy/dx = 0$ in No. 49. Next we have

$$\frac{dy}{dt} = \frac{dy}{dx} \cdot \frac{dx}{dt}.$$

Therefore

$$\frac{d}{dt}\left(\frac{dy}{dt}\right) = \frac{dx}{dt} \cdot \frac{d}{dt}\left(\frac{dy}{dx}\right) + \frac{dy}{dx} \cdot \frac{d}{dt}\left(\frac{dx}{dt}\right).$$

But the second term on the right vanishes since dy/dx is zero. Also

$$\frac{d}{dt}\left(\frac{dy}{dx}\right) = \frac{d}{dx}\left(\frac{dy}{dx}\right) \cdot \frac{dx}{dt}.$$

We have, therefore,

$$\begin{aligned} \text{acceleration} = \frac{d^2y}{dt^2} &= \frac{d^2y}{dx^2} \cdot (dx/dt)^2 \\ &= \frac{d^2y}{dx^2} \cdot v^2 \\ &= v^2/r \qquad \text{by } \textbf{No. 58}. \end{aligned}$$

It may be noted that this is a case in which there would be an advantage in using one of the former notations rather than the Leibnizian. For instance, using the dot-notation for t-derivatives and the D notation for x-derivatives, the above argument becomes

$$\begin{aligned} \dot{y}^1 &= \mathrm{D}(y) \,.\, \dot{x} \\ \dot{y}^1 &= \dot{x} \,.\, \mathrm{D}^2(y) \,.\, \dot{x} + \mathrm{D}(y) \,.\, \ddot{x} \\ &= (\dot{x})^{2!} \,.\, \mathrm{D}^2(y) \\ &= v^2/r. \end{aligned}$$

SECTION IX.

STATISTICS.

THE EXERCISES OF SECTION IX.

*** The numbers in ordinary type refer to the pages of *Exercises in Algebra*, Part II; the numbers in heavy type to the pages of this book.

EXERCISES		PAGES
CXIX.	Frequency-Distribution	433, 569
CXX.	Frequency-Curves	440, 574
CXXI.	Dispersion	449, 580
CXXII.	The Determination of Frequency by Calculation	461, 584
CXXIII.	Probability	474, 589
CXXIV.	Correlation	487, 602
CXXV.	Supplementary Examples—	
	A. Permutations and Combinations .	501, 594
	B. The Binomial Coefficients . . .	502, 595
	C. Probability	503, 595
	D. The Normal Curve	506, 596
	E. Sampling	508, 596
	F. Correlation	510, 608
	G. Partial Correlation	511, 608

CHAPTER LIII.

FREQUENCY-DISTRIBUTION.

§ 1. *Introductory.*—Statistics constitute at once the oldest and the newest branch of mathematics: the oldest, for their practice, in some form, is one of the primary necessities of ordered social life; the newest, for their theory is, to a large extent, a production of the present generation. For both these reasons it is greatly to be desired that an elementary study of the subject should come to be regarded as part of the normal programme of secondary school mathematics. On the one hand, the economic and social uses of statistics—which go back, far beyond the Conqueror's Domesday Book, to prehistoric "numberings of the people" on the banks of the Nile or the Euphrates—have shown, in recent days, wide and striking extensions which every educated person ought in some measure to understand. On the other hand, modern statisticians, seeking the most effective means of applying their weapons, have built up a striking and beautiful system of mathematical ideas. The result is that "statistics" need no longer be regarded as a synonym for "dulness"; there are few branches of mathematics which have so much that is attractive to offer to the beginner. The field has, in fact, become so rich that the task of selection for the purposes of a non-technical course is more than ordinarily difficult. The author has, however, endeavoured to limit his topics to those which are of fundamental importance and to develop them just sufficiently to give to the general body of students a fair idea of the scope and value of the subject, and to the future specialist, whether in economics, the "higher commerce," or the sciences which use statistical methods, a useful introduction to the technical methods employed in those departments.

The problems to be considered are indicated in ch. XXXVIII., § 7, and, in more detail, in the student's Introduction (p. 431). The teacher who is interested in the subject and discovers

what excellent material it offers for mathematical study may easily extend their range. With this purpose in view he will naturally turn first to Mr. Udney Yule's *Introduction to the Theory of Statistics* (Griffin & Co.), an admirable text-book to which the author is much indebted. Mr. Yule gives references to all the works which it is important for the serious student to consult. Among these the present writer has drawn most inspiration from Prof. Karl Pearson's masterly memoir (" Contributions to the Mathematical Theory of Evolution ") in the *Philosophical Transactions*, and from his other works, among which the *Grammar of Science* is widely known ; from the psychological papers of Prof. C. Spearman who has developed another side of the subject with great insight and skill ; and from some of the writings of Prof. F. Y. Edgeworth, whose articles in the *Encyclopædia Britannica* and delightful addresses to the Statistical Society (particularly, perhaps, his Presidential Address for 1912-13) are highly suggestive and stimulating. It should be added that Mr. Palin Elderton has written a book on *Frequency Curves and Correlation* (Layton) which contains a clear and useful account of the mathematical methods developed by Prof. Pearson in his Royal Society papers, and that Dr. William Brown's little book on *Mental Measurement* (Cambridge Univ. Press) is an eminently modern summary, full of valuable material not easily accessible to the general student of mathematics. *The Elements of Statistical Method* (The Macmillan Co.) by Mr. W. I. King is a semi-popular introduction which many readers would find helpful.

§ 2. Ex. CXIX. *Frequency-Distribution.*—This exercise is (i) to introduce the fundamental ideas of a frequency-distribution and its representation by the frequency-table and the frequency-diagram ; (ii) to familiarize the student with the main forms of frequency-distribution met with in naturally occurring statistics, and (iii) to supply materials to be used further in the subsequent exercises. In view of (iii) it is important that the graphs should be marked with reference numbers and preserved. It is, perhaps, inadvisable that every student should do each example ; it will, however, be well for individual students to trace, through thin paper, copies of their diagrams so that each member of the class may have a complete collection.

With reference to (i) it should be made clear how a frequency-diagram differs from the ordinary graphs of algebra or physics. These represent the relations between two variables, or the values of a function which correspond to different values of a single variable. The purpose of a frequency-diagram is simply to show how often each value of the variable is met with in a given field. Thus if, in the course of a morning's walk, I count the number of persons I meet who have black hair, brown hair, etc., I have the materials for a frequency-diagram in which the variable is the colour of the hair. The only difference between such a diagram and those to be drawn in Ex. CXIX lies in the fact that in the latter, the variable is not merely qualitative but is capable of quantitative definition. In either case any "law" to which the graphic representation may point will be simply a law of connexion or proportion among the numbers of times each value of the variable is encountered. This notion is of fundamental importance and must be thoroughly understood.

With reference to (ii), the examples give typical instances of the forms of frequency-distribution which occur most widely in statistical practice. It will be seen (1) that they are drawn from very diverse fields—from anthropometric and biological measurements, from economics, from meteorology, from physics, from medical records, and from records of the workings of what, in our ignorance, we call pure chance; (2) that, nevertheless, they display resemblances that are often most pronounced where the diversity of origin would seem to be greatest. These resemblances are brought out by the "frequency-curves" which the student is, in most cases, instructed to add to his graph. It should here be noted that the smooth curve is not, as in former uses, to preserve scrupulously the original area of the columns across which it passes. It is to represent the student's interpretation of what may metaphorically be called the *intention* of the distribution; that is, the ideal distribution to which actual samples might be expected to approach if they contained a sufficient number of cases drawn from a field sufficiently wide to be really representative. When frequency-curves are drawn with this idea in view, they are found, as was said above, to exemplify a few forms which constantly recur. The student will learn in the next exercise that these forms

have been classified by Prof. Karl Pearson under seven distinct types, of which two are symmetrical and the rest asymmetrical or "skew". The data for Nos. 4, 5, 10-14, 16 are taken from his papers on "Skew Frequency Distribution" in *Phil. Trans.*, in vols. 185, 186, 191. No. 6 is from Westergaard's *Die Grundzüge der Theorie der Statistik;* No. 7 was suggested by Prof. Edgeworth's Presidential Address referred to above; No. 8 is taken from a well-known paper in the *Philosophical Magazine;* No. 15 is drawn from Mr. Latter's article in *Biometrika*, Vol. IV.

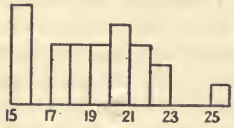


FIG. 142.

FIG. 143.

With regard to details. Fig. 142 is the histogram for the lengths of the carrots of **Nos. 1-3** when the class-interval is 2 cms., fig. 143 when it is 3 cms. It is obvious that the

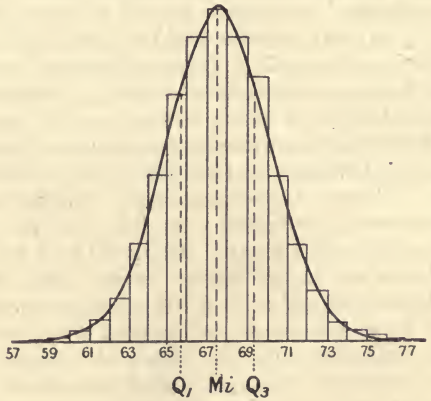


FIG. 144.

latter brings out a "law" which is quite obscured in the former. Fig. 144 shows both the actual histogram and the ideal frequency-curve of **No. 4.** It is of great importance, for it is an almost perfect instance of the "normal distribu-

tion" which has played so large a part in the evolution of statistical theory. The curve in **No. 5** is essentially of the same type but obscured by a certain amount of "skewness". **Nos. 6** and **7** are also, fundamentally, examples of normal distribution, though the symmetry of the ideal curve is in each case obscured by accidental irregularities due to the smallness of the number of cases. They are especially interesting since they exemplify the fact that pure chance, working impartially in a given set of materials, produces typically a normal distribution. This fact first came out in the study (by Laplace, Legendre, Gauss, Bravais, etc.) of the distribution of chance errors of observation. For this historical reason the normal curve is still very commonly spoken of as the graphic expression of the "law of error," even when it is actually representing things (such as statures) which are not, to the plain mind, errors at all. It is, however, useful to note that the course of statistical theory has been profoundly affected by the historical accident that it began with the doctrine of errors of observation in astronomy and surveying. Thus when the Belgian Quetelet (*c.* 1840) began the modern study of anthropometric and meteorological distributions, he fell naturally into the way of thinking of the mean of the distribution as the number (stature, chest measurement, etc.) which represents Nature's aim or intention, and deviations therefrom as her "errors". The main effect of this way of looking at natural statistics has been to prejudice inquirers unduly in favour of the normal distribution. Only recently has it been recognized that, as a matter of fact, ordinary distributions are not, even ideally, normal, and that skewness is an essential feature in them. This last point is well brought out by the distributions given in the present exercise.

In many cases, perhaps most, the asymmetry or skewness is moderate, as in the case of the head-breadths of **No. 5** or the barometer-readings of **No. 13.** Fig. 153 (p. 587) may be regarded as showing the frequency-curve typical in such cases. **Nos. 8, 12** exhibit a much more markedly skew distribution with a characteristic sharp rise on one side and an essential "tail" on the other. It is remarkable that fig. 145 represents almost equally well the distribution of scarlet fever in an epidemic and the way in which "alpha-particles" are expelled from a radio-active metal, and it is difficult to resist

the suggestion that the two sets of phenomena, though so different in outward character, are the expression of essentially similar conditions.

The extremity of asymmetry is shown by the distributions of Nos. 9, 10, and 11. Fig. 146, which actually represents the data of **No. 11**, would do, with little modification, for each of the others. This distribution is typical of many biological phenomena; infant mortality and the incidence of the death duties would be further instances.

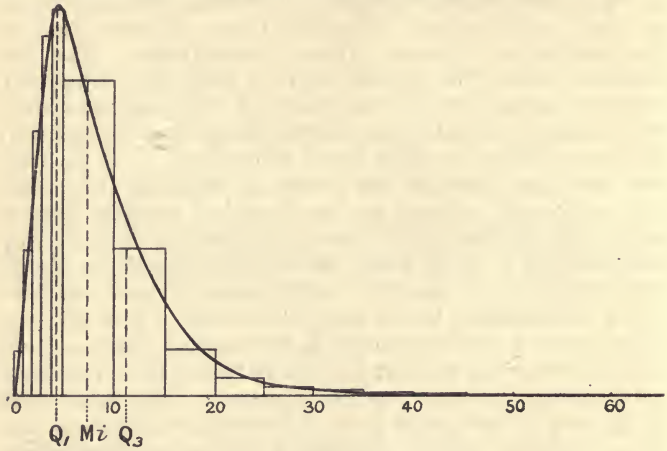


FIG. 145.

No. 14 illustrates the rather rare "**U**-distribution" in which the greatest frequencies occur at the two ends instead of towards the middle.

The curves in **Nos. 15, 16** appear to be almost perfectly normal. That of **No. 16** is the firm line in fig. 151 (p. 580). It is shown in Ex. CXX, No. 32, to be really the sum of two normal components. That of **No. 15** is probably the resultant of a still more complex system (see *Exercises*, II, p. 500).

§ 3. Ex. CXX. *Frequency-curves.*—Divisions A, B need little comment. The mode (a term introduced into statistics by Pearson) is a novelty, but the median and mean have been treated before (see pp. 45, 115, 366, and *Exercises*, I, Exs. XXVI, D; XXVII, B, C; LXVI).

In **Nos. 3, 4** the median and quartiles are determined as

follows in accordance with the investigation of **No. 2**. Half the total number of cases ($= \frac{1}{2}N$) is 4292·5 and Σn as far as the class 66-7 includes 3589 cases. In the class 67-8 there must be, then, 4292·5 − 3589 = 703·5 cases below the median. Since this class contains 1329 cases we have for the position of the median

$$Mi = 67 + 1 \times \frac{703{\cdot}5}{1329}$$
$$= 67{\cdot}52.$$

For the lower quartile we have $\frac{1}{4}N = 2146$ and $\Sigma n = 1376$ cases down to the class 64-5, leaving 2146 − 1376 = 770 cases

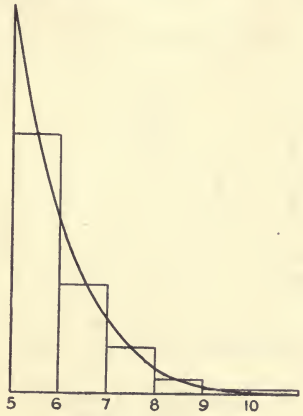


FIG. 146.

to be taken from the 990 in the class 65-6. Thus we have

$$Q_1 = 65 + 1 \times \frac{770}{990}$$
$$= 65{\cdot}78.$$

The position of the third quartile Q_3 is found in the same way, by working from the other end of the table. The positions of medians, quartiles, and means are marked in figs. 144-5.

In finding the mean the procedure illustrated in the following example (in answer to **No. 9**) should be followed.

The first column, headed X, gives the mid-values of the classes into which the lengths of the 24 carrots were grouped

in Ex. CXIX, No. 1. The column headed n states the frequency, i.e. the number of carrots in each class. By guess, the mean is about 19·5; the next two columns contain, therefore, the values of x_1, the deviations of X from this number (which is conveniently symbolized as M_1). It is best to record x_1 in two columns headed respectively + and −. The next two contain the products nx_1, also separated according to sign. The last (single) column contains the values of nx_1^2; these are not needed at present but will be required in Ex. XXII.

X.	n.	x_1.		nx_1.		nx_1^2.
		+	−	+	−	
15·5	5		4		20	80
17·5	3		2		6	12
18·5	3		1		3	3
19·5	3		0		0	0
20·5	4	1		4		4
21·5	3	2		6		12
22·5	2	3		6		18
25·5	1	6		6		36
	24			22	29	165

The algebraic sum $\Sigma(nx_1) = +22 - 29 = -7$; hence we have

$$\begin{aligned} M &= M_1 + \Sigma(nx_1)/N \\ &= 19{\cdot}5 - 7/24 \\ &= 19{\cdot}2. \end{aligned}$$

In division C the student faces the question of representing by formulæ the graphic forms which characterize so persistently statistics drawn from the most varied sources. For the present he is told simply to verify the correspondence of certain given formulæ with his frequency-curves; the derivation of these curves from definite presuppositions is to be referred to later—in Ex. CXXII and the next chapter of this book. All the formulæ are taken from Prof. Pearson's papers and represent the simpler of his "types". We start with the symmetrical curve given by the formula of **No. 17**. The various cases set in **No. 18** are represented in fig. 147; the curves marked A, B, C correspond to the examples (i), (ii),

(iii) respectively. In **No. 19** it is clear that the substitution of $-p$ for p would invert the curve and produce a symmetrical **U**.

The derivation in **No. 20** of the normal curve $y = y_0 e^{-x^2/h^2}$ from the symmetrical curve of **No. 17** is of great importance. The algebra goes as follows :—

$$\begin{aligned} y &= \underset{a \to \infty}{Lt}\ y_0 (1 - x^2/a^2)^{pa} \\ &= \underset{a \to \infty}{Lt}\ y_0 (1 - x^2/a^2)^{a^2 p/a} \\ &= y_0 \cdot (e^{-x^2})^{p/a} \\ &= y_0 \cdot e^{-x^2/h^2} \end{aligned}$$

if we put $h^2 = a/p$. It is clear that, unless p is supposed to

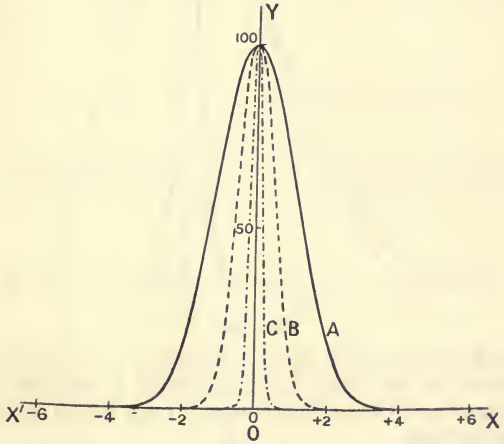


FIG. 147.

increase as well as a, the index p/a would approach zero, and the curve would degenerate into the y-axis. The normal curves to be drawn in **No. 21** are those lettered A and B in fig. 151. The ordinates may, of course, be calculated by ordinary logarithms, or, if the teacher prefers, the table of values of e^{-u^2} on *Exercises*, II, p. 473, may be used. The student must enter this table not with the value of x but with the value of x/h, i.e. $x/4{\cdot}47$ in (i) and $x/6{\cdot}33$ in (ii). The curves will be needed in No. 32.

The curve of **No. 17** is Pearson's Type II; the normal curve which we have derived from it is added by Elderton to the original list and becomes Type VII. The formula of **No. 22** is the one with which Pearson actually starts in his memoir and is, therefore, Type I. It is the curve of moderate skewness typical of so many natural distributions. The two cases of **No. 23** are shown in fig. 148, A and B.

When the indices in this formula are negative the curve is inverted and becomes the skew **U** of Ex. CXIX, **No. 14**.

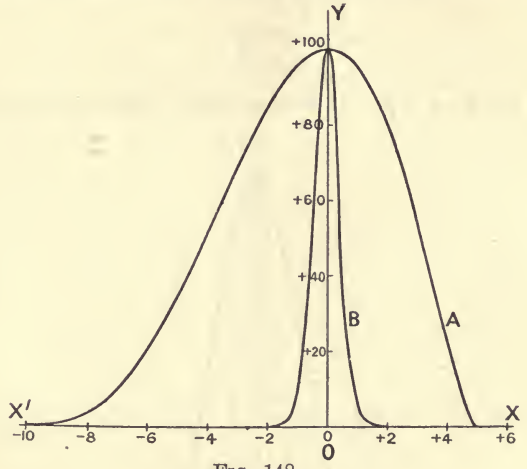


FIG. 148.

The coefficients in **No. 24** are so chosen as to give a curve very similar to the graph there obtained. It may be interesting to note that in the original memoir from which the formula is taken it reads

$$y = 50{\cdot}7505\ (1 + x/4{\cdot}8109)^{-0{\cdot}8774}\ (1 - x/5{\cdot}1705)^{-0{\cdot}943}.$$

These formidable figures will give some idea of the accuracy with which Prof. Pearson and his assistants interpret their data as well as of the incredible labour which they devote to the task.

It is easy to show in **No. 26** that the second factor in the formula of No. 22 becomes exponential when $a_2 \to \infty$. The values given for the constants are almost exactly those required to "fit" the barometric distribution of Ex. CXIX, No. 13. The curve is Pearson's Type III.

The curves of **Nos. 27-9** are of Type VI. The constants in **No. 29** will be found to give practically the curve of fig. 146.

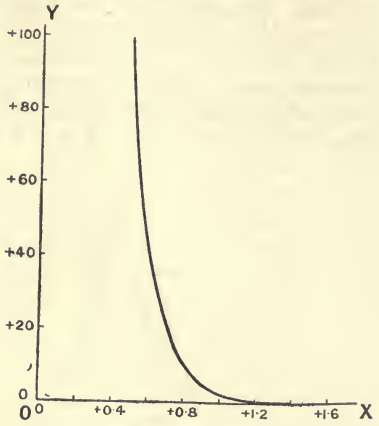


FIG. 149.

The curve given in **No. 30** is very nearly the frequency-curve of Ex. CXIX, No. 10. It is shown in fig. 149.

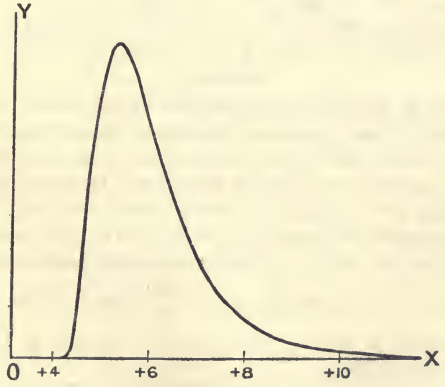


FIG. 150.

The curve of **No. 31** belongs to Type VI and is represented in fig. 150. The constants are so chosen as to reproduce very nearly the frequency-curve of Ex. CXIX, No. 12.

Lastly, in fig. 151 the curve C is the resultant of the two normal curves, A and B, drawn in No. 21 and now, in **No. 32**, to be combined. The constants are nearly those which fit the distribution of Ex. CXIX, No. 16.

§ 4. Ex. CXXI. *Dispersion.*—The notions of quartile deviation, mean deviation, and standard deviation (the last under the name "root-mean-square" deviation) have already been considered (pp. 115, 369, and Part I, Exs. XXVI, D, LXVII).

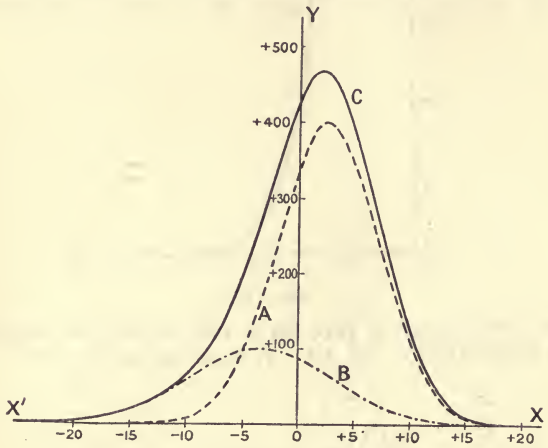


FIG. 151.

In **Nos. 3, 4** what is required is to determine an ordinate which cuts off one-quarter of the whole area of the frequency-diagram or one-half of its semi-area. Let x be its distance from the median. Then in **No. 3** we have by similar triangles that $(a - x)^2/a^2 = \frac{1}{2}$, or $x = a\,(1 - 1/\sqrt{2})$. In **No. 4** the ordinates are given by $y = b \,.\, \cos\,(\pi x/2a)$ and the area from the median up to the ordinate whose abscissa is x is

$$b \,.\, \int_0^x \cos\,(\pi x/2a) \,.\, \delta x = \frac{2ab}{\pi} \,.\, \sin\,(\pi x/2a).$$

The semi-area is $2ab/\pi$, so to find the abscissa of the quartile we have the relation

$$\sin\,(\pi x/2a) = \tfrac{1}{2} = \sin 30°.$$

Hence

$$x = 2a \,.\, \frac{30}{180} = a/3.$$

In **No. 6** we have

area of frequency-diagram = no. of cases/100.

Hence $4ab/\pi = 10$. Also $Q = a/3 = 3\frac{1}{3}$, so that $a = 10$ and $b = \pi/4$. Thus the amplitude is $\pi/4$ cms. and the base-length 20 cms.

Nos. 7-14 are of extreme importance, but do not present any particular difficulty. In **No. 7** since $x/h = u$, $x = hu$ and $\delta x = h \,.\, \delta u$. Since, also, the slices between two closely situated ordinates have the same height in each figure, the area of the one slice, $y \,.\, \delta x$, is h times that of the other slice, $y \,.\, \delta u$. Hence follows the result of **Nos. 8, 9, 10.**

The determination of the area under the normal curve is too difficult at this stage and is postponed to the Supplementary Examples. In any case the determination of the quartile deviation is possible only by some method of approximation. In the present instance, to find the whole area by Simpson's Rule we must add to half the ordinate where $u = 0$ the whole of the other ordinates of the table and multiply the sum by the common interval in u, namely 0·1. In **No. 12** the student may either verify that the area up to the ordinate $u = 0{\cdot}488$ is one-half of the semi-area under the curve, or, showing that the required abscissa is between 0·4 and 0·5, may find its length approximately by the method used to find the quartile on p. 575. In both cases it must be remembered that Simpson's Rule requires the addition of only *half* of the first and last ordinates.

The examples on mean deviation are not of great importance and could be omitted if it were desired to shorten the course. They have, however, the negative advantage of showing how good are the reasons for which statisticians prefer to use the standard deviation. The opportunity has also been taken to use them as a means of applying some of the results learnt in the parallel lessons in Section VIII. The method of calculating the mean deviation is in each case that indicated in **No. 24** in the relatively important case of a normal distribution.

On the other hand, division C, dealing with standard deviation, is of the greatest importance and should be thoroughly mastered. In **Nos. 27-8**, where standard deviations are to be calculated from a frequency-table, the work should be set down as in the table on p. 576. The last column contains the products nx_1^2, where the values of x_1 are the deviations

from the trial mean M_1. The products are obtained by multiplying the values of nx_1 in the fifth and sixth columns by the values of x_1 in the second or the third column. The total $\Sigma(nx_1^2)$ is entered at the foot of the column. We then have for the standard deviation

$$\begin{aligned} \sigma^2 &= \Sigma(nx_1^2)/N - d^2 \\ &= 165/24 - 0{\cdot}09 \\ &= 6{\cdot}78 \\ \sigma &= 2{\cdot}6 \end{aligned}$$

for in this case, as we saw on p. 576, the value of d is 0·3.

The result summed in the note after **No. 32** is one of the most frequently used in the theory of statistics and should be known by heart. The full proof of the expression for the ordinate at the origin is postponed to the Supplementary Examples. The equivalence stated in **No. 34** is also one to which reference will continually be made. The student should be quite familiar with it. The equivalence in **No. 35** is of much less importance.

CHAPTER LIV.

THE CALCULATION OF FREQUENCIES. PROBABILITY.

§ 1. *Introductory.*—Hitherto our work has been to record and analyse frequencies actually given; we are now to examine the possibility of predicting them among events that have, perhaps, never been observed. It is from this point of view and in connexion with this problem that those well-established algebraic topics—combinations, permutations, and probability—are best treated. Even if no other object were in view than to "get up" these subjects in preparing for an examination, some such method of approach as is here indicated would be the most profitable. To present the theory of combinations as a means by which the mysterious ways of chance can be foreseen and unravelled is to make an appeal to the young mathematician which he is very unlikely to resist. The teacher is, therefore, strongly advised to accompany this part of the course with select experiments intended to awaken curiosity, to supply data for investigation, and to give reality to the results of theory. By setting a class, as part of their home-work or as a voluntary "extra," the task of throwing sets of halfpennies and counting the heads, of drawing numbered cards from a hat or differently coloured marbles from a bag, of shooting at a mark with a primitive dart, etc., it is easy to collect, at very little expense in time, material of great interest and of more value than anything that can be put into a set of printed examples.

It cannot be too carefully understood and remembered that the calculation of probabilities is nothing more than the calculation of frequencies. Students are so apt to have confused ideas upon this point that, although there is no difference of principle between the examples of Ex. CXXII and those of Ex. CXXIII, yet the term "probability" has not been used in the former but is reserved until the one and only notion which the word should connote in mathematics—the notion

of "relative frequency"—has been illustrated by a considerable variety of examples.

The connexion in which the theory of combinations and permutations is here introduced has necessarily affected the character of the exercises by which it is illustrated. The investigation of these subjects leads so easily and naturally to problems of genuine importance in the theory and applications of statistics that there is little temptation to draw upon the traditional problems of the text-book. Since, however, the student who has not wrestled with these may at present be at a disadvantage in a public examination, a selection of the usual type is included among the Supplementary Examples. The teacher who is confining his attention in this section mainly to permutations and combinations should take his class straight from the present exercise to the first division of Ex. CXXV.

§ 2. Ex. CXXII. *The Calculation of Frequency.*—The key to the comprehension of this part of the subject lies in a clear understanding of the notion of "independent events". The fall of a tossed coin is an independent event; whether it will fall "head" or "tail" the next time it is thrown depends not at all upon how it fell last time or the last thousand times. For instance, if there has been a "run" of a hundred heads the "chance" that the next result will be also a head is not a whit the less than before. Assuming the coin to be unbiassed, that chance is exactly $\frac{1}{2}$—a statement which means that, if you collected a great many instances in which there had been a run of a hundred heads, "head" would be found to have been the result of the next toss also in almost exactly half of them.

The Oxford and Cambridge Boat Race may be taken as an event which is not truly "independent," in spite of a good deal of resemblance to the spinning of a coin. The victory this year is generally dependent, in part, upon what happened last year; a specially strong member of the last crew will row again, a successful "coach" will repeat his services, the previous victory will give greater confidence, etc. For these reasons the record of the Boat Race shows a large proportion of "runs" in favour of one side. Thus Oxford won continuously from 1861 to 1869 and Cambridge from 1870 to 1874.

But though a single race is not properly an independent

event yet, in the long run, the series of annual races may still present the fundamental property of independent events —namely, that, out of a great number of instances, they fall out equally often in each of the ways possible to them. As we have just seen, there are conditions which prevent the result of one race from being entirely independent of its predecessors; but these conditions may themselves be "independent" in the sense that they will, in the long run, connect themselves just as often with the fortunes of one "blue" as with those of the other. In that case the record (which now stands in favour of Oxford by 10 wins out of 68) will in time show the equality in the distribution of the results which is the typical characteristic of independent events.

The object of these remarks is to show how it is possible that events undoubtedly moulded to some extent by men "looking before and after" should yet, in the mass, conform to the same laws as the spinning of coins and the drawing of balls from urns. It is because experience shows this possibility to be, on a large scale, an actuality that it is profitable to consider in detail the dealings of the blind goddess with this kind of material.

The next idea to be grasped is that frequency-predictions are possible only in so far as the events predicted can be regarded as compounded of *independent* elementary events whose characteristic behaviour is already known. Thus, knowing that the spin of a coin is an independent event which will, in the long run, turn out heads and tails with equal frequency, we can predict with confidence what will happen (again in the long run) in the case of an event which consists in the tossing (say) of a dozen coins.

In division A of Ex. CXXII the student is confronted with simple examples to be solved directly by the principles to which reference has here been made. In **No. 8** he is called upon to formulate the theorem which is the foundation of all the more complicated predictions of the theory of probability. **No. 9** is meant to prepare the way for the theory of combinations. It is evident (i) that the two good jumps in succession may occupy four different places in the series of five, (ii) that, since each jump may be good or bad, the sum of the possibilities is 2^5, and (iii) that the required relative frequency is, therefore, $4/2^5$ or $1/8$.

In division B we turn to the theory of the combinations of

independent events. This subject should certainly be illustrated by experiments of the kind suggested in § 1. Apart from the fact that the whole discussion is made to revolve round the question of predicting frequencies instead of merely enumerating combinations, there is nothing that calls for comment until **No. 13** is reached. That example must be regarded as of fundamental importance; for, as will be shown in the next exercise, the properties of a normal distribution can be deduced from it. In the example itself, as in the complementary examples, **Nos. 15, 16**, there is no difficulty. To take the general case presented in **No. 16**: the relative frequency of a compound event in which r *specified* constituents turn out one way while the other $n - r$ turn out the opposite way is (by No. 8) $p^r \cdot q^{n-r}$, where p and q are the relative frequencies of occurrence of those two ways when the simple event is considered by itself. Now, out of the series of n constituent simple events, r can be selected to be "successes" (and the remaining $n - r$ to be "failures") in ${}_nC_r$ ways. Thus the total relative frequency of the event in question is ${}_nC_r \cdot p^r \cdot q^{n-r}$; that is, it is a definite term in the expansion of the binomial $(p + q)^n$.

It is important to note that from the condition $p + q = 1$ it follows that the total sum of the relative frequencies is (as it should be) unity. This condition, applied to the frequency-graph, implies that the total area under the curve is unity.

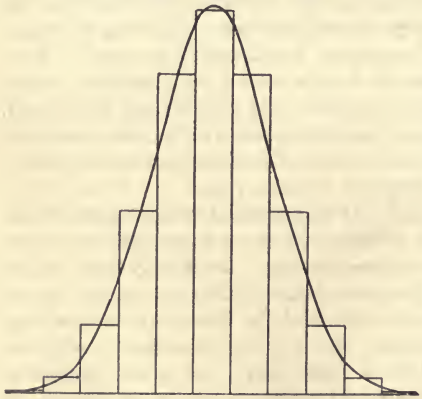


Fig. 152.

Fig. 152 shows the frequency-diagram of **No. 14.** It is evident that the smooth curve to which the distribution points as its ideal has all the appearance of being identical with the normal curve. It is worth while to give to a specially skilful and painstaking pupil the task of representing graphically

the terms of the binomial $(\frac{1}{2} + \frac{1}{2})^n$ when n has a much larger value than ten—for example, thirty. The results should be graphed, not as a series of rectangles, but as a "point-binomial" as described in Ex. CXXII, C. The curve should be drawn through the points and the normal curve added in which $\sigma^2 = \frac{1}{4}(n + 1)c^2$ where c is the interval between the abscissæ of the points. The approximation of the two curves will be found very close.

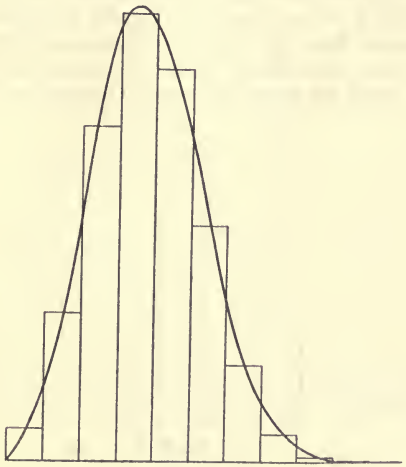


Fig. 153.

Fig. 153 shows the graph of **No. 14.** It evidently points to a "moderately skew" frequency-curve just as the former distribution points to the normal curve. Here, again, it would be profitable to have several curves drawn, varied values being assigned to p and q but always so that $p + q$ is unity. It is interesting to note that the distribution represented by the point-binomial

$$(0{\cdot}9 + 0{\cdot}1)^{30}$$

(the coefficients for which may have already been calculated for the normal curve) is a fairly close approximation to the distribution of barometric heights given in Ex. CXIX, No. 13.

The derivation of the normal curve from the point-binomial is the subject of division C. The subject is of great import-

ance but the exposition should occasion little difficulty. The teacher will note the justification for the assumption that w^2n is finite although w^2x is negligible; it is merely another way of saying that, however large x may be taken, the curve must be supposed to extend so far beyond the point in question that the ratio x/wn is negligible.

Another mode of deriving the normal curve from the point-binomial is given in § 5 of this chapter.

In division D we turn from combinations to permutations. There is nothing in the text or examples to attract attention until we reach **No. 34.** This is intended to prepare the student for the interesting result discussed in § 6 of the exercise—a result for which the name "Spearman's Theorem"

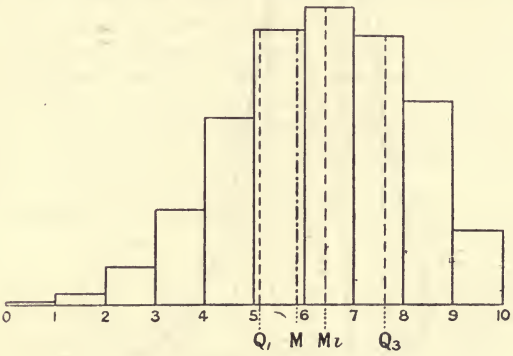


FIG. 154.

is proposed on account of its connexion with Prof. C. Spearman's "foot-rule" method of measuring correlation to be described in Ex. CXXIV. Spearman's own proof is given as an appendix to a paper in the *British Journal of Psychology*, Vol. II, pt. i. (1906).

The frequency-distribution for the total "loss of rank" when, as in **No. 40,** $n = 6$ is shown in fig. 154. It will be seen that, although the diagram is not symmetrical as a whole, yet the central part does not depart very far from symmetry while the left-hand half resembles that of a normal distribution. Both these features become more marked as n increases.

The theorem lends itself to easy and attractive experimental illustration—illustration which will be found very

useful in preparation for the discussion of correlation in Ex. CXXIV. In an experiment performed by the author seventeen small and equal cards were labelled A, B, . . . Q, were shaken together in a hat, and drawn at random, the order of appearance of each letter being noted and the total loss of rank being estimated as described in the text. The process was repeated 100 times. Fig. 155 is the frequency-diagram for the different values of L thus obtained. It differs from the diagram obtained theoretically for $n = 6$ in

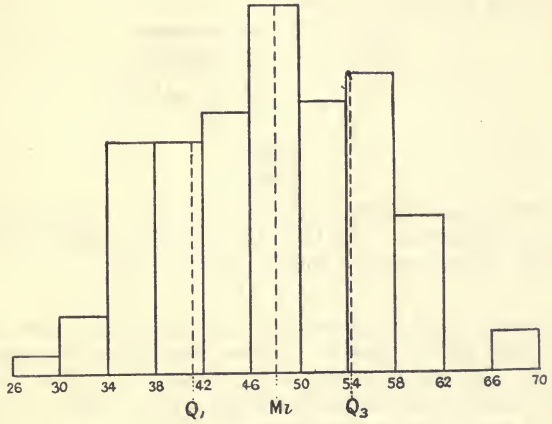


FIG. 155.

that only the central part of the whole theoretical distribution appeared in the experiment. By theory the average value of L ($= \bar{L}$) should have been $(17^2 - 1)/6 = 48$; the actual averages were as follows:—

After	10	20	30	40	50	60	70	80	90	100 drawings.
$\bar{L}$	48·5	48·2	48·9	48·5	47·7	47·9	48·3	47·4	47·2	47·5

§ 3. Ex. CXXIII. *Probability.*—Division A introduces formally the term probability and gives simple examples to illustrate its use. Some are based on calculated, some on observed, frequency-distributions. **No. 5** (ii) is a typical instance of the former kind. Each spin of the tee-to-tum gives three possible events, two "blues" and one "red". The total number of possible cases for ten spins is, therefore,

3^{10}. The number of cases in which blue and red are each obtained five times is

$$_{10}C_5 \cdot 2^5.$$

For, if you had ten tee-to-tums before you, you could choose five of them to be "blue" in $_{10}C_5$ ways; and, taking any one of the selected tee-to-tums you could make it give "blue" in two ways, so that any selection of five gives 2^5 different cases of "all blue". By a similar argument "three blues only" is obtainable in $_{10}C_3 \cdot 2^3$ ways. Thus, by the alternative definition given in § 1 of the text

$$\text{prob.} = \frac{\text{favourable cases}}{\text{all cases}}$$

$$= \frac{_{10}C_5 \cdot 2^5 + _{10}C_3 \cdot 2^3}{3^{10}}$$

$$= \frac{3008}{19683}$$

The same result could, of course, have been obtained directly from the formula of Ex. CXXII, No. 15, by adding together the two corresponding terms of the expansion

$$(\tfrac{2}{3} + \tfrac{1}{3})^{10};$$

for probability and relative frequency are synonymous terms.

Of the second class of problem **No. 6** may be taken as typical. By the frequency-table of Ex. CXIX, No. 4, out of 8585 British adults 336 are six feet high or more. The required probability is, therefore, 336/8585.

In **No. 9** we turn from the use of the frequency-table to that of the frequency-diagram or curve. Here we have to remember simply that the area between any two ordinates measures the number of cases whose magnitudes fall within the limits indicated by the abscissæ of those ordinates and that the whole area of the diagram measures the total number of cases. Hence the formula of No. 9. When the diagram represents not absolute but relative frequencies its total area A is, of course, unity.

Nos. 11, 12 give an extremely important instance of the application of this simple principle, and the main results of the table in **No. 12** will be used constantly. Their importance consists in the fact that the relations which hold good exactly for normal distribution hold good approximately for many other distributions. Thus the great bulk of most

distributions is found, in practice, to lie within the range 6Q on either side of the median, and about 96 per cent of it within the range 3Q. The student should check this statement roughly by reference to his frequency-diagrams, not failing to note some (e.g. that of Ex. CXIX, No. 14) in which it fails entirely. No doubt there has been a tendency among statisticians to use the principle too freely—a tendency derived, once more, from the mistaken supposition that the normal is also actually the usual form of natural distributions. Still, used cautiously, the principle is of great service.

In division B we turn to the calculation of "compound probabilities". Here the fundamental principle of Ex. CXXII, No. 8, is used in a more elaborate way for the solution of more difficult problems, but there is no new principle to be learnt. **Nos. 13, 14** give a slight indication of the way in which the theory of probability is applied by actuaries to the calculation of premiums for benefits conditioned in more complicated ways than the ordinary single life assurance. (The teacher should, by the way, take a suitable opportunity of pointing out that, in Exs. LXXX, LXXXI, the use made of the Life Table was a simple anticipation of these calculations of probability; it is easy to modify the arguments there used so as to bring in formally the notion of relative frequency.) In **No. 14** the probability that the company will have to pay the £200 is the sum of the probabilities (i) that both brothers will be alive in forty-four years' time, (ii) that brother A will be alive and brother B dead, (iii) that brother B will be alive and brother A dead; for in all these contingencies the money must be paid. There remains only one possible case, namely, that both are dead. In that case the company will not have to pay the benefit. This is the easier probability to compute. From the Life Table (*Exercises*, II, p. 73) it will be seen that the probability of A's death before sixty-five is

$$\frac{70846 - 33234}{70846} = 0{\cdot}53076.$$

The similar probability for B is $(1 - 33234/71780) = 0{\cdot}53687$. The probability that both will be dead is the product of these single probabilities, i.e. 0·28495, and the probability that one at least will be alive is, therefore, $1 - 0{\cdot}28495 = 0{\cdot}71505$. If P is the premium paid in precisely similar

circumstances by (say) 100,000 uncles, the sum to be distributed by the company at the end of 44 years will be £100,000P × 3·67 and the number of sums of £200 each to be paid out of it will be 71,505. Hence

$$P = £200 \times 0{\cdot}71505/3{\cdot}67$$
$$= £39 \text{ nearly}$$

From this example it is easy to deduce the rule followed by actuaries in computing the premium for any given benefit: premium = (present value of benefit) × (prob. that it will be paid)

Nos. 16-18 are important for the reasons indicated in the note that follows them and expanded in § 5 of this chapter. To take the typical term of the general case, it is clear that if, of the n balls, r are black and the rest white, the number of possibilities of r black is ${}_{pm}C_r$ and of $n - r$ white ${}_{qm}C_{n-r}$. Thus the total number of "favourable cases" or "successes" is the product of these numbers. But the "all possible cases" are evidently ${}_mC_n$; hence the probability of the event is

$${}_{pm}C_r \cdot {}_{qm}C_{n-r} / {}_mC_n.$$

It is easily verified that the successive terms of the hypergeometric series give the values of this probability as r assumes the values $n, n-1, n-2, \ldots, 0$. Fig. 156 exhibits the frequency-diagram in the case of **No. 17**.

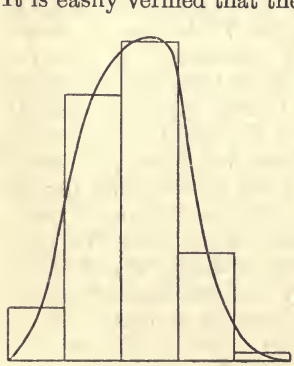


Fig. 156.

With division C we approach some of the most useful and attractive uses of the theory of probability outside actuarial practice. The exposition given in the text is full enough to make further explanation unnecessary. The data for **Nos. 21, 22** are taken from Merriman's well-known book, *The Method of Least Squares*, to which the reader should refer for full information about the reduction of observations. Gauss's classical treatise (which was translated into French by Bertrand) will also be found very illuminating and not very difficult. The subject is too technical for treatment in a general course.

Merriman gives the following for the theoretical numbers

in **No. 21**, basing them upon a formula slightly different from the one in the text: 3, 15, 50, 118, 197, 234, 197, 118, etc. For **No. 22** he gives (quoting Bessel): 107, 87, 57, 30, 13, 5, 1, 0, 0, 0. It is to be noted that Bessel does not give the signs of his errors; we must assume, therefore, that positive and negative errors of the same magnitude occurred with equal frequency.

Here again is a very profitable field for original experiment. The author has obtained very fair results by the following two methods: (i) A dart was aimed a large number of times at a point marked on a sheet of paper hung upon a door. Concentric circles of radii 1 in., 2 in., 3 in., etc., were drawn on the sheet and the number of shots falling within each band counted. The usual assumption was made that all the shots in a band lay upon its mid-line, the standard deviation was calculated and the frequency formula deduced as in No. 21. (ii) A line 10 cms. long was drawn upon a postcard. The experimenter looked at it and then marked off on one of the blue lines of a sheet of exercise paper, provided with a red marginal line, the length which he judged to be 10 cms. He covered over his first attempt, refreshed his memory of the original line and made a second attempt. In this way a large number of judgments of one individual may be obtained and the frequency-formula deduced. Other experiments with the same object will readily suggest themselves.

Careful attention should be given to the beautiful theorem of the "standard error" of the mean in § 4. It is important that the correct meaning of the unfortunate term "probable error" should be appreciated. In this connexion the note in **No. 24** may be useful. In the proof of the theorem itself the only difficulty likely to be felt lies in the assumption that $\Sigma(a_r a_s)$ is zero. When Ex. CXXIV has been read the student will see that this is merely the assumption that the different deviations that enter into a group of observations are not "correlated" with one another in pairs; that is, that the entry of one into a given group is quite independent of the entry of any other. In Ex. CXXIV it will be proved formally that in that case $\Sigma(a_r a_s)$ tends, as the number of cases increases, to become zero.

In **No. 25** we have, for the standard error of the mean height of 25 men, $\sigma = 2{\cdot}57/5 = 0{\cdot}514$. With the usual assumption that different deviations of the mean of a sample

from the mean of the whole follow a normal distribution, we deduce that the probable error is $0{\cdot}534 \times 0{\cdot}675 = 0{\cdot}36$. In **No. 26** the given mean of $68\frac{1}{2}$ in. represents a deviation from the general mean of $68{\cdot}5 - 67{\cdot}46 = 1{\cdot}04$ in. Since $Q = 0{\cdot}36$ this deviation is very nearly $3Q$. By reference to the table appended to No. 12 on p. 476 we see that the probability of so great a deviation is about 1/24. It may be expected, that is to say, once in each 20 to 25 batches of men.

An interesting experiment upon this subject will be to take a table of logarithms, or square roots, etc., and select a number of groups of last figures at random. For example, out of the last figures of the 900 columns of an ordinary 4-place or 5-place table of logarithms or antilogarithms, groups of 9 may be selected by one student, groups of 16 by a second, groups of 25 by a third. The mode of selection does not matter provided that it is truly a random selection and gives each part of the table the same chance. Since the digits run from 0 to 9 and occur (no doubt) with practically equal frequency among the original 900, we may take 4·5 as the mean value of a digit. The sum of the squares of the deviations of these 900 digits from zero would be $100(0^2 + 1^2 + \ldots + 9^2) = 28500$. The mean square of the deviation from zero is, therefore, 285/9 and the square of the deviation from 4·5 is $285/9 - 81/4 = 11{\cdot}42$. Thus for the whole collection of digits $\sigma = 3{\cdot}38$. In accordance with the theorem the standard deviations of the different means obtained by random selection of groups of 9, 16, 25 digits should be 3·38/3, 3·38/4, etc. It will be understood that the closeness of the experimental to the theoretical numbers depends upon the largeness of the number of selections. The theory supposes an exhaustive number.

The "standard error of sampling," the subject of division D, is also an extremely important matter that lends itself easily to experimental illustration. The teacher should not fail to have simple experiments carried out in accordance with the suggestions given by the examples here and in Ex. CXXV.

§ 4. Ex. CXXV. *Supplementary Examples.*—The first five groups of the supplementary examples may most profitably be considered in this chapter. Division A has already been mentioned; it consists of additional problems in permutations and combinations, all of a conventional character and of types familiar from the text-books. The teacher who wishes to de-

velop this subject further should consult the second volume of Chrystal's *Algebra* or the well-known *Choice and Chance* of W. A. Whitworth. The same remarks also apply to the examples in division B. They are included almost solely in view of the examinations in which they are occasionally set.

In division C a few examples are given upon the subject of "local probability". The assumption underlying the solution of these problems is that a line or an area can be regarded as made up of a very large number of equidistant points. This assumption is, of course, arbitrary. For example, if the number of points between any two given points on the line is infinite, the distribution of probabilities of which the final ar-

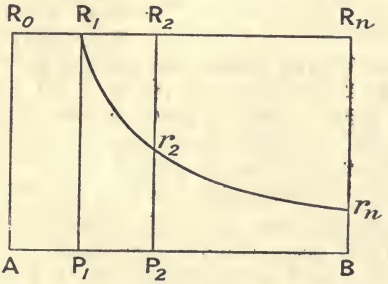


FIG. 157.

rangement is the limit need not be uniform at all. Thence, as the reader who has looked into the subject of probability knows, many "paradoxes" have arisen. The simple examples of Ex. CXXV, C, avoid these and they must be sought in special books of which Todhunter's *History of the Theory of Probability*, Williamson's *Integral Calculus*, Borel's *Eléments de la Théorie des Probabilités* (Paris, Herman), and Czuber's *Wahrscheinlichkeitsrechnung* (Leipzig, Teubner), are, perhaps, the best for the purpose.

The double frequency-diagram required for the solution of **No. 24** is shown in fig. 157. The rectangle gives, as explained in the text, "all possible cases" of the double event: "first break at P, the second anywhere on the broken piece". The area $AR_0R_1r_2r_nB$ gives the "favourable cases". This area is composed of two parts; the area of the rectangle R_0P_1 is b (since $AR_0 = 1$) and that under the curve is the integral

of $b \cdot \delta x/x$ between the limits $x = b$ and $x = a$. The result given in the text follows at once.[1]

Fig. 158 is the double frequency-diagram by which the student is directed in **No. 25** to solve Buffon's famous problem. "All cases" are represented by the rectangle whose area is $2a$. The "successes" are represented by the area under the curve, which is the integral of arc cos $(x/a) \cdot \delta x$ from $x = 0$ to $x = b$. The ratio of the two areas is $2a/\pi b$. Many persons have sought to "determine π experimentally" by using this result, a rod (e.g. an uncut lead pencil) being thrown a large number of times upon a ruled floor and a record being kept of the proportionate number of times it falls across a line.[2]

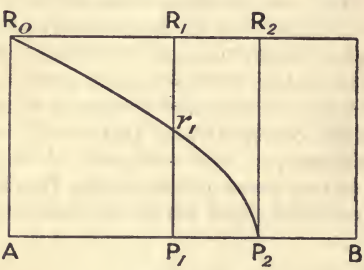


FIG. 158.

In division D the investigation of Ex. CXXI, C, is completed by a proof that the area under the normal curve $y = e^{-x^2}$ is $\sqrt{\pi}$. The proof given was in part suggested by a question set by Drs. Bromwich and Forsyth in the B.A. Mathematical Honours examination of the University of London in 1912. From the result it follows by Ex. CXXI, No. 9, that the area of $y = y_0 e^{-x^2/2\sigma^2}$ is $y_0 \cdot \sigma\sqrt{(2\pi)}$. If, then, the whole area under the curve is to be equal to the number of cases whose frequency-distribution it represents we must have

$$y_0 \cdot \sigma\sqrt{(2\pi)} = \mathrm{N}$$

or

$$y_0 = \mathrm{N}/\sigma\sqrt{(2\pi)}$$

as we have so often assumed to be the case.

Division E gives further examples of the most interesting and important subject of sampling errors. The quotations from Prof. Bowley are made in part from his Presidential

[1] The author has to acknowledge the kindness of Mr. C. S. Jackson who pointed out the source of a subtle paradox in which he became involved in a first attempt to solve this problem by a simple method. The problem itself is taken from Williamson's *Integral Calculus;* No. 23, from Borel.

[2] Mr. B. Branford informs the author that his mathematical class obtained a very satisfactory approximation by this method.

Address to the Economics Section of the British Association reported in the *Journal of the Royal Statistical Society* for 1906, partly from his *Elementary Manual of Statistics*. As Prof. Bowley points out, the great importance of the theory of sampling is in connexion with inquiries into unemployment, wages, the cost of family life, etc., which should embrace the whole population in their scope. It is obvious that it is impossible in such cases actually to review the whole population ; we are necessarily driven back upon sampling. The theory becomes then of immense importance as a means of estimating the value of the results obtained from a given sample. It shows that, provided the sample is selected truly at random from the field of the whole population, so that every member has an equal chance of being taken, the conclusions drawn will apply with surprisingly little modification to the whole. The difficulty will generally be in securing a really random selection of the sample. To select the companies for the investigation described in **Nos. 32-3** Prof. Bowley numbered all in the list; he then read down the last four figures of a table in the *Nautical Almanack* and whenever he came upon a number less than 3878 took the corresponding company to be an element of the random sample.

With the aid of a collection of statistics, such as those contained in *Whitaker's Almanack*, the *Statistical Abstract* or the *Statesman's Year-book*, the teacher will easily be able to supply a number of interesting examples of sampling of the same type. Examples similar to **No. 31** can also be contrived with little difficulty.

The proof of the interesting theorem in **No. 37** is very simple. Since $q = 1 - p$ we have, when p is small, $pq = p - p^2 = p$ approximately. Hence the standard error for the number of successes becomes simply $\sqrt{(pn)}$. But pn is, by hypothesis, the number of successes S. Hence the standard error is $\sqrt{S}$. The important point to note here is that, if p is small, the standard error can be determined without any knowledge of the actual ratio of successes to the whole. Thus in **No. 39**, if we had no other information than that 119 railway passengers were killed in 1912 and that the number is a very small proportion of the travelling public, we could still obtain a fair value for the standard error. It will be $\sqrt{119} = 10{\cdot}9$ and the probable error $10{\cdot}9 \times 0{\cdot}675 = 7{\cdot}35$. Since $156 - 119 = 37$, i.e. rather more than five times the probable

error, a mortality roll of 160 passengers might (by the table on p. 476) be expected to occur (with our present population and conditions of traffic) not oftener than once in about three thousand years.

If the teacher is a student of physics he will be especially interested to note, in connexion with **No. 40**, that the standard deviation in the rate at which the atoms of radio-active metals discharge "alpha-particles" brings the phenomenon into line with events like fatal railway accidents. No doubt the discharge of an "alpha-particle" is, like death in a railway collision, an incident of comparatively great rareness in the society in which it occurs. We have here a striking modern instance of the well-known principle that it is possible by statistical theory to give an account of many physical phenomena, the kinetic theory of gases being the most notable form in which that principle has hitherto expressed itself. In his Presidential Address to the Royal Statistical Society (published in the Society's *Journal* for 1912) Prof. Edgeworth has dwelt in a peculiarly illuminating way upon this aspect of statistical science.

§ 5. *The Derivation of Frequency-curves.*—This chapter may conveniently end with a brief sketch of the method by which Prof. Karl Pearson derives his different types of frequency-curves.

Consider, in the first place, the point binomial $(\frac{1}{2} + \frac{1}{2})^n$ and the normal curve $y = y_0 e^{-x^2/2\sigma^2}$. Take any two adjacent points, Q_1, Q_2, on the former and find the point P on the latter where the tangent is parallel to the line Q_1Q_2. Also let PM be the ordinate of P. Then the analogue of PM in the point-binomial is neither the ordinate at Q_1 nor that at Q_2 but the ordinate QN which is drawn from Q, the mid-point of Q_1Q_2. Thus $QN = \frac{1}{2}(y_r + y_{r+1})$, y_r being the term ${}_nC_{r-1} \cdot (\frac{1}{2})^n$ in the binomial expansion. If, further, we put $x_r = rc$ where c is the distance between the ordinates of successive points of the figure, we have the relation

$$\frac{(y_{r+1} - y_r)/c}{\frac{1}{2}(y_{r+1} + y_r)} = \frac{c(n+2) - (x_{r+1} + x_r)}{\frac{1}{2}(n+1)c^2}.$$

Now take as y-axis a line $\frac{1}{2}(n+2)c$ to the right of the present origin; then the new abscissa $x'_r = x_r - \frac{1}{2}(n+2)c$. Also put $\frac{1}{2}(n+2)c^2 = 2\sigma^2$. Then the denominator of the second fraction becomes $-(x'_{r+1} + x'_r)$ and the above relation can be written

$$\frac{\text{slope of } Q_1 Q_2}{\text{mean ordinate}} = -\frac{\text{mean abscissa}}{\sigma^2}.$$

But in the normal curve we have the relation

$$\frac{1}{y} \cdot \frac{dy}{dx} = -\frac{x}{\sigma^2}.$$

Thus there is between the normal curve and the polygon formed by joining the points of the point binomial a geometrical similarity that is quite independent of the value of n.

By a similar argument we may derive in the case of the asymmetrical point binomial $(p + q)^n$ a relation of the form

$$\frac{\text{slope of } Q_1 Q_2}{\text{mean ordinate}} = -\frac{X'_{r+\frac{1}{2}}}{a + X'_{r+\frac{1}{2}}}$$

where $X'_{r+\frac{1}{2}}$ is the mean abscissa measured from a convenient point. Hence it is seen that the skew point-binomial is geometrically similar to the curve

$$y = y_0 (1 + x/a)^{\gamma a} \cdot e^{-\gamma x}$$

for the differential equation of this curve is

$$\frac{1}{y} \cdot \frac{dy}{dx} = -\frac{x}{a + x}.$$

But the point-binomial cannot be made to supply analogues for frequency-curves of all the types actually encountered. For this reason Prof. Pearson turns to the analogy of the urn, described in Ex. CXXIII, B, and seeks the differential equation which shall correspond to the relation between the slope of the line Q_1Q_2 joining the points given by the two consecutive terms of the hypergeometric series to the ordinate mid-way between them. It is not difficult to obtain the relation in question in the form

$$\frac{\text{slope of } Q_1 Q_2}{\text{mean ordinate}} = \frac{x}{b_0 + b_1 x + b_2 x^2}$$

x being the mean abscissa measured from an appropriate origin. For the corresponding continuous curve we have, then, the differential equation

$$\frac{1}{y} \cdot \frac{dy}{dx} = \frac{x}{b_0 + b_1 x + b_2 x^2}.$$

By giving different values to the constants the different types may be obtained. Four of them have been set, with numerical coefficients, as No. 26 (vi) – (x) of Ex. CXVIII. Those omitted can be derived without difficulty from the former.

§ 6. *The "Fitting" of Frequency-curves.*—Supposing a

given distribution to be recognized as of a certain type there still remains the problem of finding for the formula the coefficients which will make the curve "fit" the data as closely as possible. Prof. Pearson's method may be made clear in principle by considering the way in which the formulæ were derived in Ex. CXIII, Nos. 21, 22, where it was assumed that the type was the normal distribution. In this case there were two constants to determine—the constant which fixes the origin and the standard deviation. The first is found by applying the familiar condition that if the mean is taken as the origin then Σnx is zero, n being the number of cases whose magnitude is x. In other words, if the mean does not in the first instance coincide with the origin the corresponding constant in the formula will be derived from the value of Σnx. Similarly the second constant, the standard deviation, is found from the value of Σnx^2. Now it is obvious that in this case *all* the data are taken into account in determining the values of the two constants, so that the resulting curve may fairly be claimed to be the one which represents best their general voice. In the case of a skew curve these two sums do not suffice to give the constants. Prof. Pearson accordingly suggested that the values of Σnx^3, Σnx^4, etc., should be used to find them; for each of those sums will be related to some integral derived from the formula of the curve in the same way as the standard deviation is related to the integral $\int x^2 y \,.\, \delta x$.

By analogy with mechanics Prof. Pearson calls the sums Σnx, Σnx^2, etc., the first, second, etc., "moments" of the data. The calculation of the constants by their means is, as may be supposed, a tedious business and far too complicated for full discussion here. Reference should be made either to Prof. Pearson's memoirs already quoted or to Mr. Elderton's more accessible book on *Frequency Curves* in which the method of moments is explained thoroughly and applied to examples. Meanwhile the following simple investigation may serve to exemplify its general character.

Let it be known that a given frequency-distribution is of the type whose "ideal" is described by a differential equation of the form

$$\frac{dy}{dx} = \frac{yx}{a + bx}$$

then our task is to find for the constants a and b values which

will give due weight to all the measurements contained in the distribution. Write the equation in the form

$$(a + bx)y' = yx$$

and multiply each side by $x^n \,.\, \delta x$. We thus obtain the relation

$$x^n(a + bx)y' \,.\, \delta x = x^{n+1}y \,.\, \delta x$$

or

$$ax^n y' \,.\, \delta x + bx^{n+1}y' \,.\, \delta x = x^{n+1}y \,.\, \delta x.$$

Next integrate the three products by the ordinary rule for "integration by parts," taking as limits the values of x where the curve cuts the x-axis. In this way we obtain

$$(ax^n + bx^{n+1})y\Big]_{x_1}^{x_2} - a\int_{x_1}^{x_2} x^{n-1}y \,.\, \delta x - b\int_{x_1}^{x_2} x^n y \,.\, \delta x = \int_{x_1}^{x_2} x^{n+1}y \,.\, \delta x.$$

But, by hypothesis, y is zero for both the limits, so the first term disappears. Also the integrals that remain are, by definition, the $(n - 1)$th, nth and $(n + 1)$th moments of the distribution about the vertical throughout the origin. Thus our result may be compactly expressed by the notation

$$-a\mu_{n-1} - b\mu_n = \mu_{n+1}.$$

Now let $n = 1$; then μ_0 becomes $\int y \,.\, \delta x$, i.e. the total area under the curve. If the numerical data are *relative* frequencies the area is unity. Also, if the origin s taken at the mean value of the variable, the first moment $\mu_1 = \int xy \,.\, \delta x$, is zero. Thus the equation reduces to

$$a = -\mu_2.$$

Similarly, if $n = 2$, since μ_1 is zero we have

$$b = -\mu_3$$

whence it follows that the differential equation of the frequency-curve is

$$\frac{dy}{dx} = -\frac{x}{\mu_2 + \mu_3 x}.$$

Finally the constants are to be given numerical form by the relations $\mu_2 = \Sigma(nx^2)$ and $\mu_3 = \Sigma(nx^3)$, where the values of x are those given in the frequency-table.

CHAPTER LV.

CORRELATION.

§ 1. *Introductory.*—The measurement of correlation is the latest, and in some ways the most important of the achievements of modern statistical science. For the person who is seeking to straighten out the tangle of facts and relationships which meets and baffles the inquirer into heredity, social and economic phenomena and psychology—to say nothing of meteorology—the correlation coefficient is proving an instrument of indispensable usefulness and unsuspected power. Our course may appropriately conclude with a study of the ideas underlying its theory and some simple examples of its application. For further study the student must turn where his special interests draw him: to the memoirs of Prof. Pearson and his school if it be to biology and "eugenics," to the writings of Mr. Yule and others if it be to economics, to the papers of Prof. Spearman and his pupils, of Dr. William Brown and other Englishmen or to those of Prof. Thorndike and other distinguished Americans if it be to psychology. In all these directions he will find that the introduction of the quantitative measurement of correlation is justifying old Roger Bacon's prescient dictum that "mathematics is the gate-way to *all* the sciences".

§ 2. Ex. CXXIV. *Spearman's Coefficient.*—The choice of a means of measuring correlation is to some extent arbitrary and it is well that the student should see it to be so. That is one reason for the study of Prof. Spearman's method before the "standard" method is considered. A second reason is that it is extremely simple to understand and to apply, and is therefore of great use as an introduction to the general notion of the measurement of correlation. It should be noted that (as is implied by the name "foot-rule") its inventor does not regard it as a rival to the standard method,

but rather as a quick and ready means of determining whether there is in a given instance a degree of correlation which would warrant the application of further labour to the interpretation of the data.

Another advantage which the method possesses, from the teaching point of view, is the readiness with which original material can be found in schools to illustrate its use. Investigations of the correlation between the performances of a class in different subjects, in the same subject in different terms, in different examinations in the same subject, in school performances which are not both "subjects," etc. : all these would be useful from the mathematical standpoint and would often give information of no mean value to the teacher and so to his pupils. Particularly important in this connexion is the use of the correlation coefficient as a measure of the "reliability" of an examination or other test.

The student is not likely to fail to understand the exposition of division A in the light of the investigation of Ex. CXXII, E. The greatest difficulty, the necessity for a knowledge of the probable error, has also been anticipated in Ex. CXXIII. The calculation in **Nos. 4-12** of Spearman's formula for this constant is not very elegant; his own (given as an appendix to the paper already quoted from the *British Journal of Psychology*, Vol. II) is too difficult for reproduction here. Since, however, it is impossible to give here the proof for the probable error of the standard coefficient it seemed advisable to find a simple, if indirect, method of establishing the formula in the present case.

The illustrative example in § 2 comes from the paper just referred to. **No. 1** is taken from a paper in the *Journal of Psychology* for December, 1911. The data for **No. 16** are given by Prof. Spearman in his well-known paper on "General Intelligence" in the *American Journal of Psychology* for April, 1904. The data of **No. 17** were kindly placed at the author's disposal, together with a full calculation of the related correlation coefficients, by Miss N. Carey.

§ 3. Ex. CXXIV, B, C. *The Standard Method.*—The origin of the standard method of measuring correlation is to be found in a remarkable paper by the French *savant* Bravais published in 1846. Bravais's problem was to determine the most probable position of a point when the errors in its coordinates are not unconnected with one another. To solve

the problem it was necessary to obtain some measure of the degree of connexion between the errors in x and y. In this way Bravais was lead to the "product-moment" which, in the hands of Prof. Pearson, has become the standard means of measuring correlation.

Pearson, like Bravais and Galton who preceded him, worked at first with the assumption that the distribution of the correlated variables was in each case normal. Upon this assumption Galton's "regression lines" would actually pass through the means of the various arrays. The wider interpretation of the regression-line given in the text is due to Mr. Yule. Galton's choice of the term "regression" has reference to the fact, which his graphic method brings out so well, that (for example) the mean height of the sons of a very tall or a very short man is nearer than his own height to the "mediocrity" of the general mean.

The late Prof. Weldon invented, in illustration of the Pearson method of measuring correlation, experiments of the following character. (The data and some details of the experiment are taken from an interesting paper by Mr. A. D. Darbyshire in the *Memoirs* of the Lit. and Phil. Soc. of Manchester for 1907.) Take a dozen dice, and make in parallel columns a record of the scores of a large number of *pairs* of throws. It is evident that between the two sets of scores thus obtained there would be little or no correlation, for there is no connexion between the two members of a pair of scores except that which may be due to chance. The case is like the drawing of labelled cards from a hat as described on p. 589. But now mark a certain proportion of the dice by dyeing them in red ink, and after throwing the whole set and taking the first score, leave the red dice on the table and throw the residue again. To make the second score add to the number now obtained from these the number shown by the red dice lying on the table. A treatment, by the Bravais-Pearson method, of the pairs of scores thus obtained should now reveal a distinct correlation; for the second score of a pair depends in part upon the first score in every case. In one of the experiments performed for Mr. Darbyshire, out of twelve dice four were dyed red. Five hundred pairs of throws were made with the results shown in the following table. This may be regarded as a typical "correlation table" showing in clear form the vertical and

horizontal arrays whose means, in the ideal case, lie in or near the regression lines. The vertical arrays may be regarded as first throws, the horizontal as second throws.

	0	1	2	3	4	5	6	7	8	9	10	11	12	
0														
1					1	1								2
2			1	1	1	2	1	1						7
3			2	4	3	4	5	2						20
4			1	2	8	18	8	12	6	1	1			57
5			3	8	16	16	19	21	7	5	2			97
6		1	1	5	19	25	25	20	12	2				110
7			2	1	6	22	17	32	17	12	3			112
8					5	6	16	18	14	7	2			68
9						3	2	6	5	2				18
10							3		2	2				7
11								1	1					2
12														
	0	1	10	21	59	97	96	113	64	31	8	0	0	500

Fig. 159 presents the results of this experiment graphically, the small circles indicating the means of the vertical arrays, the crosses those of the horizontal arrays. To find b_1 and b_2, the gradients of the regression lines, it is necessary to determine first the means of the whole of the first throws as set out in the bottom line and of the whole of the second throws as set out in the right-hand line of the correlation table. The coefficients must then be calculated by the formulæ

$$b_1 = \Sigma(xy)/\Sigma x^2$$

and $b_2 = \Sigma(xy)/\Sigma y^2$, x and y being respectively the deviations of the two numbers of a pair of scores from the means of the totals.

Let X be the value of a first throw and Y that of a second throw. Then to find $\overline{X}$ the numbers at the head of each array must be multiplied by the partial total at the foot of that array, the products summed, and the sum divided by 500:

$$(1 \cdot 1 + 2 \cdot 10 + 3 \cdot 21 + \ldots + 10 \cdot 8)/500 = 6{\cdot}086.$$

Similarly to find $\overline{Y}$ we have

$$(1 \cdot 2 + 2 \cdot 7 + 3 \cdot 20 + \ldots + 11 \cdot 2)/500 = 6{\cdot}062$$

the multipliers being taken from the last column.

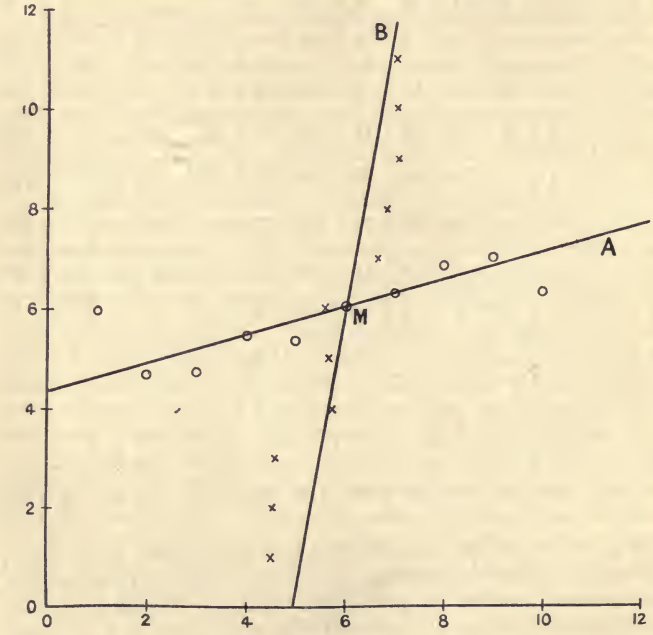


FIG. 159.

To find the mean of the vertical array which sets out all the second scores that were obtained with a first score of (say) 4, the numbers in that array must be multiplied, each by the number at the beginning of its row, and the sum of these products divided by the number at the foot of the array: $(1 \cdot 1 + 2 \cdot 1 + 3 \cdot 3 + 4 \cdot 8 + \ldots + 8 \cdot 5)/59$. Similarly to find the mean for the horizontal array which

records the values of all the first scores associated with a *second* score of 4, each number in the array must be multiplied by the number at the head of its column, and the sum of the products divided by the number at the end of the array: $(2 . 1 + 3 . 2 + 4 . 8 + \ldots + 10 . 1)/57$.

To calculate b_1, b_2, and r we must know the values of $\Sigma(xy)$, $\Sigma(x^2)$, and $\Sigma(y^2)$. In practice it will be better to determine these from the values of $\Sigma(XY)$, $\Sigma(X^2)$, and $\Sigma(Y^2)$. We know that

$$\Sigma(x^2) = \Sigma(X^2) - N\overline{X} \text{ and } \Sigma(y^2) = \Sigma(Y^2) - N\overline{Y}.$$

To determine $\Sigma(xy)$ from $\Sigma(XY)$ we have the relation

$$\begin{aligned}\Sigma(XY) &= \Sigma(\overline{X} + x)\,(\overline{Y} + y) \\ &= N\overline{XY} + \Sigma(xy)\end{aligned}$$

since Σx and Σy are both zero.

In finding $\Sigma(XY)$ the sixteen cases whose scores are recorded in the table in heavy type will contribute $4 \times 5 \times 16$, the 4 and 5 being the values of x and y and the 16 the number of associations of those values. Similarly the twelve cases recorded in heavy type contribute $9 \times 7 \times 12$ to the total sum. In this way $\Sigma(XY)$ will be found to be 18,835. ΣX^2 will be found by multiplying the squares of the numbers at the top of the table, each by the corresponding total recorded at the bottom: $1^2 . 1 + 2^2 . 10 + 3^2 . 21 + \ldots = 19{,}900$. In the same way

$\Sigma(Y^2) = 1^2 . 2 + 2^2 . 7 + 3^2 . 20 + \ldots + 11^2 . 2 = 20{,}747$

the multipliers being the totals in the final column.

In addition to these numbers we have $\overline{X}^2 = 37{\cdot}038$, $\overline{Y}^2 = 36{\cdot}74$, and $\overline{XY} = 36{\cdot}88$. Hence, by the above relations it follows that

$$\Sigma(xy) = 394,\ \Sigma(x^2) = 1381,\ \Sigma(y^2) = 2373$$

whence $b_1 = 0{\cdot}286$, $b_2 = 0{\cdot}166$, and $r = \sqrt{(b_1 . b_2)} = 0{\cdot}218$.

When (as is usually the case) correlation is to be calculated without a correlation table from a pair of records such as those headed C_a and C_b in **No. 17**, the best mode of procedure is indicated by the specimen on the next page.[1]

The first two columns contain the data. The mean for the first column is 43, that for the second 40. The two columns headed "x" contain the difference between the mean and the values of D_a, positive differences being separated from negative. The differences for the D_b data are entered

[1] Copied from Miss Carey's notebook, with alterations.

similarly in the next two columns. The squares follow and finally the products. The last three columns being summed give all the information required for the calculation of b_1 and b_2 (if they are needed) and of r.

The products xy can be obtained (when they are too difficult to be found mentally) by reference to a multiplication table. Thorndike's *Mental and Social Measurements* contains such a table up to 100 × 100. For elementary purposes, however, a table is unnecessary. An alternative is to add columns giving the values of $(x + y)$, $(x + y)^2$, $(x - y)$, and $(x - y)^2$, the squares being taken from a table of squares,

D_a.	D_b.	x.		y.		x^2.	y^2.	xy.	
		+	−	+	−			+	−
40	40		3		0	9		0	
37	43		6	3		36	9		18
53	46	10		6		100	36	60	
37	30		6		10	36	100	60	
61	55	18		15		324	225	270	
34	30		9		10	81	100	90	

which is a commoner thing than a table of products. The subtraction of $(x - y)^2$ from $(x + y)^2$ gives $4xy$. $\Sigma(xy)$ is then obtained by summing the $4xy$ column and dividing the result by 4.

§ 4. Ex. CXXV, F, G. *Supplementary Examples.*—The last two divisions of the supplementary exercise deal with further important matters connected with the measurement of correlation. The first is the means by which faulty data may be used in combination to give a truer verdict than they could give alone. Here we follow Prof. Spearman whose studies of this important subject have been especially fruitful. The proof given is, however, essentially that of Mr. Yule. Prof. Spearman's proof is given (in its more developed form) in the *Journal of Psychology* for October, 1910. Lastly, comes a short division devoted to the subject of Partial Correlation. This vitally important contribution to the theory and practice of correlation-measurements is due to Mr. Yule. He has treated it in a very general way in his *Introduction*, illustrating his exposition by an appeal to an instructive model.

Consideration is restricted here to the case of three variables; it is, however, easily seen that the method can be extended to any number. The researches of psychologists and statistical economists will offer many instances of the use of Yule's formula, of which that given in **No. 50** may be regarded as typical.

INDEX.

Adams, M., 453.
Addition, algebraic, 54, 164, 184.
Aim in teaching mathematics, 16.
Algebra, nature of, 1; essentials of course in, 51.
Angles; *see* Sine and Cosine, Tangent; complementary, 128; in spherical triangles, 454; of unlimited magnitude, 60, 394, 501.
Annuities, 432.
Antilogarithms, 306, 331, 333, 339; antilogarithmic function, 310, 342; antilogarithmic curve, 345; antilogarithm of an irrational, 432.
Approximation-formulæ $(a \pm b)^2$, $(a \pm b)^3$, 72.
Approximations; *see* Calculus.
Archimedes, 169, 178, 256, 257, 292.
Area, calculation of, by methods of calculus, 169, 203; directed areas, 207; area-functions, 251, 279; area of surface of sphere, 443.
Argand, J. R., 390, 470.
Arithmetic, and algebra, 1; generalized, 25; of infinites, 170.
Arithmetical series, 168, 199.
Association, law of, 193.
Astronomy, 387, 457.

Ball, R., 459, 462.
Ball, W. W. R., 24.
Barnard, S., 60.
Bessel, 593; functions, 562.
Bidder, G., 373.
Binomial theorem, 211, 372.
Board of Education, 78.
Bonola, R., 452.
Boole, Mary, 39.
Borel, E., 595, 596.
Bosanquet, B., 4.
Bowley, A. L., 596.
Bradley, F. H., 3.
Branford, B., 24, 373, 596.
Bravais, A., 573, 603, 604.
Briggs, H., 57, 311, 438.
Bromwich, T. J. I'A., 596.
Brouncker (Lord), 438.
Brown, W., 570.
Buffon, G. L. Le C., 596.

Cajori, F., 24.
Calculus; position of the subject, 19; calculation of areas and volumes by constant difference series, 169, 203; Wallis's Law, ordinate and area-functions, 246, 251, 279; differential formulæ, 282; generalization of Wallis's Law, 374; differential formulæ of the exponential and logarithmic functions, 435; differential formulæ with polar coordinates, 504; differential formulæ of the circular functions, 528; differentiation by the method of limits, 549; D notation, 549; Leibnizian, 549, 560, 566; dot notation, 566; integration by the method of limits, 553; $\int$ notation, 554; Leibnizian, 553; differential equations, 556, 560; partial differentiation, 563; total differentials, 563; curvature, 565.
Cantor, G., 17, 384, 404, 405, 407, 408, 410.
Carson, G. St. L., 24, 162.
Cauchy, A. L., 559.
Cavalieri, A., 169, 170.
Cayley, A., 474, 493.

Centroids, 367.
Changing the subject of a formula, 77, 78, 104.
Child, J. M., 60.
Chrystal, G., 397, 495, 595.
Circular functions, 295, 499, 512; inverse, 514.
Circular measure, 60, 499.
Clifford, W. K., 452, 476.
Column-graphs, 36, 571.
Combinations, 60, 397, 399, 583, 594.
Commensurable magnitudes, 411.
Commutation, law of, 197.
Complex numbers, 390, 391, 469; products of, 475; complex and real variables, 478; relations between complex variables, 486; logarithm of a complex number, 493.
Component and resultant, 185.
Compound interest, 347.
Constant-difference series, 168, 199.
Constant-ratio series, 176, 224.
Constants in a formula, determination of, 136.
Continuum, number of the, 408.
Contour lines, 426.
Correlation, 398, 589, 602; Spearman's coefficient, 602; Bravais-Pearson coefficient, 603; correlation table, 605; partial correlation, 608.
Couturat, L., 404.
Cosine; *see* Sine and Cosine.
Curriculum in mathematics, 17.
Curvature, 565.
Curve of pursuit, 38.
Curve of squares, 149.
"Cut" (*Schnitt*), 414.
Cycloids, 506.
Czuber, E., 595.

Darbyshire, A. D., 604.
Darwin, G. H., 520, 527.
Davis, R. F., 148, 150.
Dedekind, R., 404, 405, 413, 416.
De Moivre, A., theorem of, 390, 391, 469, 472, 476.
Descartes, R., 390, 469.
Differential calculus; *see* Calculus.
Differential formulæ; *see* Calculus.
Directed areas, 207.
Directed numbers, 54, 159, 162, 181, 183, 228, 230; addition and subtraction, 164; multiplication and division, 168, 193.
Division, algebraic, 176, 193, 222.
Dunlop, H. C., 306.

e, 58, 300, 311, 350, 418, 434, 483, 495, 534.
Edgeworth, F. Y., 570, 572, 598.
Edser, E., 311.
Elderton, P., 570, 578, 600.
Ellipse, transformations of, 363; elliptic functions, 532.
Epicycloids, 507.
Equality, sign of, 8.
Equations, simple, 77-9, 113; simultaneous, 233; quadratic, 80, 238, 270; further, 240; differential, 248.
Error, law of, 593; standard error, 593; probable error, 593; probable error in correlation, 603.
Euclid, 4, 32, 70, 411.
Examinations, public, 19, 60, 382, 398.
Expansions, 72, 73, 174, 177, 372.
Exponential curve, 345, 433.
Exponential function, 343, 428, 433.
Exponential values of the sine and cosine, 494; of the hyperbolic sine and cosine, 534.

Factorization, 68, 75, 82, 209.
Fahrenheit, D. G., 228.
Fluxions, 425, 488.
Formulæ, 6, 25, 63, 96; substitution in, 30, 67; changing the subject of, 77, 78, 104; combining of, 113, 136; generalization of, 119, 146; relation to graphs, 119; differential formulæ, 435.
Forsyth, A. R., 596.
Fourier, J. B. J., 393, 516; his theorem, 520, 563.
Fractional indices, 281, 343.
Fractions, algebraic, 13, 75, 96, 222, 224, 231.
Frequency: distributions, 398, 569; curves, 574; dispersion, 580; determination of, by calculation,

583; derivation of frequency curves, 598; fitting of *ditto*, 600; relative frequency, 584.
Functions, 46, 110; inverse, 243, 274; linear, 110, 235, 258, 259; hyperbolic and parabolic, 236, 264; area, 279; ordinate, 280; periodic, 392; continuous and discontinuous, 422; of two variables, 426; singular values, 426.

Galton, F., 604.
Gauss, C. F., 17, 390, 470, 474, 479, 573, 592.
Generalization, types of, 2.
Generalized arithmetic, 25.
Geometric series, 176, 224.
Girard, A., 390.
Gnomonic nets, 447.
Goodwill, G., 237, 476.
Gosiewski, L., 558.
Gradient, 34, 252, 376, 425.
Granville, W. A., 556.
Gray, A., 562.
Graphs, 31, 418, 422, 486, 504, 507, 509, 513; interpolation, 34; column-graphs, 36, 571; principles of method, 40; graphs in practical work, 43; medians and quartiles, 45; graphs and co-ordinate geometry, 47; three-dimensional graphs, 47, 479; practical suggestions, 48; relation to formulæ, 31, 67, 119; *see also* Hyperbola, Parabola, Straight Line.
Great circle sailing, 447.
Gregory, J., 438; his series, 393, 514, 529, 530.
Growth problems, 58, 302, 313; growth curves, 304, 317, 319; growth-difference, 315; growth-factor, 313; growth-factors, nominal and effective, 346.
Gudermannian functions, 536.
Gunter, E., 57, 301, 321.
Gunter-scale, 302, 304, 319, 326, 333.

Hall, Stanley, 303.
Hamilton, W. R., 476.
Hankel, H., 475.
Hardy, A. S., 470.
Hardy, G. H., 423, 424, 494, 495, 537.
Harmonic analysis, 520.
Harmonic motion, simple, 501, 517; compound, 509, 517.
Hayward, R. B., 476.
Heath, T. L., 32, 257, 411.
Henrici, O., 476.
Hermite, C., 419.
Hinks, A. R., 442.
Hipparchus, 256, 295.
History of mathematics, 4, 9, 14, 16, 24, 32, 46, 57, 58, 70, 77, 169, 212, 256, 281, 292, 300, 301, 302, 311, 321, 372, 376, 390, 404, 411, 419, 469, 569, 595.
Hobson, E. W., 421.
Hoefler, A., 24.
Hulbert, L. S., 556.
Huntingdon, E. V., 310.
Hutton, C., 257, 530.
Hyperbola, as symbol of inverse proportion, 145; hyperbolic functions, 236, 264, 531; rectangular, 147, 264; transformations of, 363; differentials of hyperbolic functions, 535.

i, 391, 469, 475, 480, 486.
Imaginary numbers, 60, 238, 390, 391, 398, 469, 484.
Incommensurable magnitudes, 411.
Index notation for numbers, 214.
Indices, integral, 174; fractional, 281, 343; positive, 214; negative, 217.
Induction, 282.
"Infinitesimals," 553.
Insurance, life, 432, 591.
Integers; *see* Number Systems.
Integral calculus; *see* Calculus.
Interest, compound, 347.
Inverse functions, 243, 274.
Inverse proportion curve, 145.

Jackson, C. S., 306, 558.
James, W., 4.
Jevons, W. S., 3, 15.
Jourdain, P. E. B., 23.

Keith, T., 389.
Kelland, P., 476.

Kelvin (Lord), 527.
King, G., 433.
King, W. I., 570.

Lagrange, J. L., 17, 559.
Lambert's net, 443.
Laplace, P. S., 573.
Latitude, circles of, 129; middle latitude sailing, 129.
Latter, O., 572.
Least squares, method of, 43, 592.
Legendre, A. M., 573.
Leibniz, G. W., 549; notation for differentiation, 549, 560, 566; for integration, 553.
Limits, theory of, 171, 396, 541; practical applications, 546.
Lindemann, F., 419.
Linear functions, 110, 235, 258, 259.
Lissajou's curves, 509.
Logarithms, 57, 299, 325, 333; tables of, 329, 337; common, 309, 335, 337; Napierian, 308; change of base, 306; use of, in trigonometry, 355; logarithmic function, 310, 342; theory of logarithms stated algebraically, 341; logarithms as indices, 343; logarithm of an irrational number, 432; of a complex number, 493.
Lubbock, J., 526, 527.

Machin's series, 530.
Map projections, 386, 441.
Maseres, F., 438, 530.
Mathews, G. B., 562.
Maxwell, J. Clerk, 17, 476.
Mean deviation, 115, 366, 580.
Mean position, 366.
Mean value, theorem of, 557.
Medians, 45, 115, 366, 574.
Mercator, N., 57, 385, 438.
Mercator net, 446, 537.
Mercator sailing, 132, 445.
Meridian gnomonic net, 447.
Merriman, M., 592.
Method of differences, 269, 282.
Milborne, 302.
Milne, J. J., 148, 150.
Minus sign, meanings of, 181, 185.
Mitchell, U. G., 10.
Mode, 574.
Molesworth, G. L., 67.
Mollweide's equation, 356.
Moments, theory of, 600.
Moore, E. H., 558.
Multiplication, algebraic, 172, 193, 207.

Napier, John, 57, 58, 300, 308, 329, 355, 425, 438.
Napierian logarithms, 308.
Nautical mile, 129.
Negative numbers, 159, 183.
Newton, I., 4, 170, 372, 425, 488, 549.
Nominal and effective growth-factors, 346.
Normal curve and distribution, 572, 574, 577, 587, 598; area of curve, 581, 586, 590.
Numbers, non-directed, 52, 61; directed, 54, 159, 162, 181, 183, 228, 230; rational, 409, 413; irrational, 410, 413, 432; real, imaginary, complex, cardinal, 406; ordinal, 405; infinite, 406; algebraic, 390-1, 413, 419; transcendent, 419; *see also* Imaginary and Complex Numbers.
Number of the continuum, 408.
Number-scale, the complete, 179, 228, 421.
Number systems, 403.

One-to-one Correspondence, 405, 407, 452.
Ordinate functions, 248, 281, 480.
Oughtred, W., 10, 302.
π, 27; calculation of, 256, 292, 530; transcendental nature of, 418, 421.

Parabola, as symbol of direct proportion to the square, 149; parabolic functions, 236, 266; parabolic formula, changing subject of, 274; parabola, area of, 251, 279; transformations of, 363.
Partridge, S., 302.
Peano, G., 558.

Pearson, Karl, 370, 562, 570, 572, 574, 576, 578, 600, 604, 608.
Periodic functions, 392, 515, 531.
Permutations, 60, 397, 399, 583, 594.
Perry, J., 24, 311, 556.
Physics and mathematics, relation between, 394.
Plus sign, meanings of, 183, 185.
Polar co-ordinates, 356, 504.
Polar gnomonic net, 447.
"Point-binomial," 587, 598.
Positive and negative numbers, 54, 159, 162, 182, 228, 230.
Preston, T., 508.
Probability, 385, 397, 583, 589.
Probable error, 593; in correlation, 603.
"Product-moment," 604.
Projections, map, 386, 441.
Proportion, direct, 109, 117; inverse, 109, 145; direct proportion to the square or square root, 149; inverse proportion to the square or square root, 152; combination of types of proportion, 154.
Ptolemy, 32, 256, 295.
Pythagoras, 411.

Quadratic equations, 80, 238, 270.
Quadratic functions, 236, 266, 359.
Quartiles, 45, 574; quartile deviation, 115, 580.
Quetelet, L. A. J., 573.

Rate of change of a variable, 34, 170, 425.
Rational numbers, 409, 413.
Rayleigh (Lord), 562.
Recurrence, proof by, 282.
Regiomontanus, J., 256.
Regression, 604.
Resultant and component, 185.
Rhumb lines, 445.
Riemann, G. F. B., 392, 474, 487, 488, 503.
Rolle, M., 396.
Root-mean-square deviation, 115, 369, 580.
Roulettes, 507.
Round numbers, 214.
Russell, Bertrand, 23, 384, 404, 408, 416, 419, 420, 453, 544, 546, 550.

Sampling, 594.
Sanson's net, 442.
Sidereal time, 459.
Signs, rule of, 53, 173, 186, 197, 211.
Silberstein, L., 424, 477, 558.
Simple equations, 233.
Simpson's rule, application of, 523, 581.
Sine and cosine, 111, 124; relations between, 133; relations with tangent, 132; values for 45°, 30°, 60°, 134; directed, 255, 261; calculation of, 295; sum and difference formulæ, 296, 359, 508; exponential values of, 494; extension to angles of unlimited magnitude, 501; curves of, in wave-motion, etc., 517; differential formulæ for, 528; expansions of, 529; hyperbolic, 531.
Sinusoidal net, 442.
Slide rule, 302, 304, 328, 523.
Smith, D. E., 24, 419.
"Space-filling curve," 558.
Spearman, C., 570, 588, 602.
Sphere, area of surface of, 443; *see also* Trigonometry.
Spherical defect, 452.
Spherical excess, 451, 466.
Spherical triangles, 454.
Spirals, 504.
Square root, 68, 70, 73, 81, 90; curve, 151.
Squares, curve of, 149.
Stallo, J. B., 452.
Standard deviation, 115, 370, 580.
Standard error, 593.
Standard form of numbers, 175, 310, 328.
Statistics, 115, 214, 397, 569.
Stereographic projection, 449.
Stifel, M., 212, 270, 372, 411.
Straight line, as symbol of direct proportion, 109, 119.
Subtraction, algebraic, 54, 164, 184.
Sun-dials, 460.
Surds, 71, 93, 231.
Symbolism, 4, 9, 10, 63.

Symbols, substitutes for words, 6; manipulation of, 12; literal, introduction of, 26.

Tait, P. G., 476.
Tangent, 111, 121; of 90°, 123; tables of, 122, 256; relations with sine and cosine, 132; values for 45°, 30°, 60°, 134; directed, 236, 259; extension to angles of unlimited magnitude, 501; differential formula for, 529.
Tannery, P., 413, 416.
Taylor's Theorem, 397, 557, 559.
Theon of Alexandria, 70.
Tides, 525.
Todhunter, I., 595.
Triangles, solution of, without formulæ, 112; relations between sides and trigonometrical functions of angles, 263; spherical, 454.
Trigonometry, position of the subject, 19; trigonometrical ratios: tangent (height problems), 121; sine and cosine (navigation problems), 124; latitude and longitude, 129; relations between sine, cosine, and tangent, 132; ratios of 45°, 30°, 60°, 134; combination of formulæ, 143; the tangent, directed, 255, 259; the sine and cosine, directed, 255, 261; relations between the sides of a triangle and the ratios of the angles, 263; calculation of π, 292; calculation of sines, 295; use of logarithms in trigonometry, 355; sum and difference formulæ (angle-sum less than 360°), 359; trigonometry of the sphere, map projections and navigation problems, 386, 441, 537; trigonometry of spherical triangles, astronomical problems, 387, 454; exponential values of the sine and cosine, 493; circular functions: circular measure, angles of unlimited magnitude, spirals, roulettes, 499, etc.; harmonic motion, simple, 501, 517; compound, 509, 517; sum and difference formulæ for angles of unlimited magnitude, 508; inverse circular functions, 514; wave-motion, 515.
Tuckey, C. O., 174.
Turner, G. C., 476.
Turning values, 267.

Van Ceulen, L., 256, 292; table of, 295.
Variation, 154.
Vectors, 112, 124, 128, 165, 185, 188; algebra of, 188, 404, 476.
Venn, J., 6.
Vieta, F., 9, 12.
Volumes, calculation of, by methods of the calculus, 169, 203.
Von Wyss, C., 33.

Walker, G. T., 358.
Wallis, John, 2, 3, 9, 12, 21, 57, 169, 250, 257, 281, 367, 372, 385, 530.
Wallis's Law, 246, 279, 287, 374, 396, 436, 437, 552, 555; theorem, 559.
Waves, 515, 536.
Weierstrass, K., 396, 558.
Weldon, W. F., 604.
Wessel, C., 470.
Westergaard, H., 572.
Whewell, W., 527.
Whitehead, A. N., 5, 23, 57, 77, 545, 550.
Whitworth, W. A., 595.
Williamson, B., 595, 596.
Wingate, E., 302.

Young, J. W. A., 10, 23, 24, 160, 404, 408, 410, 419, 421, 453, 545.
Yule, G. U., 370, 570, 604, 608.

Zeno, 384; paradoxes of, 419.

ABERDEEN: THE UNIVERSITY PRESS

RETURN TO the circulation desk of any
University of California Library
or to the
NORTHERN REGIONAL LIBRARY FACILITY
Bldg. 400, Richmond Field Station
University of California
Richmond, CA 94804-4698

ALL BOOKS MAY BE RECALLED AFTER 7 DAYS

- 2-month loans may be renewed by calling (510) 642-6753
- 1-year loans may be recharged by bringing books to NRLF
- Renewals and recharges may be made 4 days prior to due date.

DUE AS STAMPED BELOW

JUL 27 1999

12,000 (11/95)

YB 35355

632257

LB1645

N8

UNIVERSITY OF CALIFORNIA LIBRARY